SVEUČILIŠTE U RIJECI
POMORSKI FAKULTET

Mario Musulin

PRILOG POVEĆANJU SIGURNOSTI OBALNE PLOVIDBE BRODOVA PRIMJENOM MODELA PROCJENE RIZIKA

DOKTORSKI RAD

Rijeka, 2026.

SVEUČILIŠTE U RIJECI
POMORSKI FAKULTET

Mario Musulin

PRILOG POVEĆANJU SIGURNOSTI OBALNE PLOVIDBE BRODOVA PRIMJENOM MODELA PROCJENE RIZIKA

DOKTORSKI RAD

Mentor: dr. sc. Đani Mohović, red.prof.
Komentor: dr. sc. David Brčić, izv. prof.

Rijeka, 2026.

UNIVERSITY OF RIJEKA
FACULTY OF MARITIME STUDIES

Mario Musulin

A CONTRIBUTION TO IMPROVING THE SAFETY OF COASTAL SHIP NAVIGATION THROUGH THE APPLICATION OF RISK ASSESSMENT MODELS

DOCTORAL THESIS

Rijeka, 2026.

Mentor doktorskog rada: dr. sc. Đani Mohović, red.prof.
Komentor doktorskog rada: dr. sc. David Brčić, izv. prof.

Doktorski rad obranjen je dana_____na Sveučilištu u Rijeci, Pomorski fakultet, pred Povjerenstvom u sastavu:

1. dr. sc. Mirano Hess, redoviti profesor u trajnom izboru Sveučilišta u Rijeci, Pomorskog fakulteta, predsjednik,
2. dr. sc. Davor Šakan, docent Sveučilišta u Rijeci, Pomorskog fakulteta, član,
3. dr. sc. Luka Mihanović, redoviti profesor u trajnom izboru Sveučilišta obrane i sigurnosti „Dr. Franjo Tuđman“, Hrvatska, član.

SAŽETAK

Pomorski promet u obalnim vodama karakterizira povećana izloženost riziku zbog složenih navigacijskih uvjeta, ograničenog prostora za manevriranje, velike gustoće prometa te promjenjivih meteoroloških i oceanografskih uvjeta. Tradicionalni pristupi procjeni rizika često su nedovoljno učinkoviti u pravovremenom i pouzdanom prepoznavanju prijetnji, ponajprije zato što ne integriraju sustavno dostupne podatke niti uzimaju u obzir dinamičku prirodu operativnog okruženja. U suvremenim analizama sigurnosti naglasak se sve više stavlja na širok skup čimbenika rizika koji obuhvaćaju uvjete na moru, karakteristike broda, obrasce plovidbe, meteorološke okolnosti te druge relevantne parametre koji mogu utjecati na nastanak pomorskih događaja.

Svrha provedenog istraživanja bila je uspostaviti metodološki okvir za primjenu strojnog učenja u analizi i procjeni rizika pomorskih nezgoda u obalnim vodama. Poseban je naglasak stavljen na znanstveno utemeljenim modelima koji omogućuju pouzdanije razumijevanje obrazaca rizika i podržavaju donošenje sigurnosno utemeljenih odluka. Istraživanjem se nastojalo unaprijediti postojeće pristupe procjeni rizika te pružiti temelje za razvoj operativno primjenjivih rješenja koja mogu podržati sigurniju i učinkovitiju plovidbu.

Istraživanje je provedeno primjenom naprednih metoda strojnog učenja, uključujući slučajne šume, potporne vektore i neuronske mreže, na podacima agencija za istraživanje pomorskih nezgoda u razdoblju 2020.–2024. Podaci su prethodno sistematizirani, očišćeni i pripremljeni za analitičku obradu, dok su modeli naknadno testirani, optimizirani i validirani kako bi se osigurala njihova prediktivna učinkovitost te sposobnost generalizacije u stvarnim operativnim uvjetima plovidbe.

Rezultati su pokazali da integrirani modeli omogućuju preciznu identifikaciju područja povišenog rizika, predviđanje nezgoda te analizu važnosti pojedinih čimbenika rizika. Na temelju usporedne analize performansi modela predložene su konkretne mjere za smanjenje rizika, uključujući unaprjeđenje navigacijskih protokola, edukaciju posada, primjenu suvremenih tehnologija i prilagodbu zakonodavnog okvira.

Istraživanjem su razvijeni i evaluirani modeli za analizu evidentiranih pomorskih događaja te metodološki okvir koji može poslužiti kao osnova za budući razvoj sustava potpore odlučivanju u obalnoj plovidbi.

Ključne riječi: *pomorski promet, obalne vode, analiza rizika od nezgoda, sigurnosni čimbenici, predikativno modeliranje, strojno učenje, slučajne šume, potporni vektori, neuronske mreže*

ABSTRACT

Maritime traffic in coastal waters is characterised by increased exposure to risk due to complex navigational conditions, limited manoeuvring space, high traffic density, and variable meteorological and oceanographic factors. Traditional risk assessment approaches are often insufficiently effective in ensuring timely and reliable threat detection, primarily because they do not systematically integrate available data or account for the dynamic nature of the operational environment. Contemporary safety analyses increasingly emphasise a broad spectrum of risk factors, including sea state conditions, vessel characteristics, navigation patterns, meteorological circumstances, and other relevant parameters that may influence the development of incident scenarios. The purpose of this research was to establish a methodological framework for applying machine learning to the analysis and assessment of maritime accident risks in coastal waters. Special emphasis was placed on scientifically grounded models that enable a more reliable understanding of risk patterns and support safety decision-making based on evidence. The study aimed to enhance existing risk assessment approaches and provide foundations for the development of operationally applicable solutions to support safer and more efficient navigation. The research employed advanced machine learning methods, including random forests, support vector machines, and neural networks, applied to datasets from maritime accident investigation agencies covering the period 2020–2024. The data were systematically organised, cleaned, and prepared for analytical processing, and the models were subsequently tested, optimised, and validated to ensure predictive performance and the ability to generalise to real operational navigation conditions. The results indicated that integrated models enabled accurate identification of high–risk areas, prediction of accidents, and assessment of the importance of individual risk factors. Based on a comparative performance analysis of the models, specific measures were proposed for risk reduction, including improvements to navigational protocols, enhanced crew training, the application of advanced technologies, and adaptations to the regulatory framework. The research developed and evaluated models for the analysis of recorded maritime events, as well as a methodological framework that can serve as a basis for the future development of a decision support system for coastal navigation.

Keywords: *maritime traffic, coastal waters, accident risk analysis, safety factors, predictive modelling, machine learning, random forests, support vector machines, neural networks*

PREDGOVOR

Iznimna je čast završiti ovaj doktorski rad, koji nije samo rezultat mojeg individualnog truda, već i podrške, znanja i ohrabrenja koje sam primao tijekom cijelog istraživačkog procesa. Stoga osjećam duboku potrebu izraziti zahvalnost onima koji su svojim doprinosom oblikovali ovaj doktorski rad i moje akademsko putovanje.

Prije svega, iskreno zahvaljujem svom mentoru, dr. sc. Đaniju Mohoviću, red.prof., na njegovom neprocjenjivom znanju, stručnom vodstvu i dugogodišnjem iskustvu koje mi je bilo čvrst oslonac u svim fazama istraživanja. Njegova posvećenost, razumijevanje i povjerenje značajno su utjecali na kvalitetu ovog rada, a njegova spremnost da podijeli svoje uvide i uvijek pronađe vremena za savjet ostavit će trajan trag u mom profesionalnom razvoju.

S posebnom zahvalnošću ističem i svog komentora, dr. sc. Davida Brčića, izv. prof., čija su analitičnost, sistematičnost i konstruktivne primjedbe znatno doprinijele oblikovanju sadržaja rada. Njegova preciznost u izražavanju, sposobnost da prepozna ključna pitanja i uputi na suštinske smjernice, predstavljali su važan doprinos akademskoj jasnoći i kvaliteti ovog istraživanja.

Također, zahvaljujem članovima Povjerenstva na njihovom stručnom doprinosu, kritičkom pregledu i korisnim sugestijama koje su omogućile dodatno unaprjeđenje rada. Njihov angažman i interes za temu dali su ovom istraživanju dodatnu vrijednost.

Posebnu zahvalnost upućujem pomorcima, studentima i stručnjacima iz prakse čija su iskustva, promišljanja i uvidi poslužili kao vrijedna inspiracija te doprinijeli boljem razumijevanju teorijskih i praktičnih aspekata obalne plovidbe. Njihovo aktivno uključivanje bilo je ključno za povezivanje teorijskih spoznaja s realnim izazovima obalne plovidbe.

Neizmjerno sam zahvalan svojim prijateljima i kolegama koji su, bilo razgovorom, savjetom ili ohrabrenjem, doprinijeli tome da ovaj izazovan put bude lakši i ispunjen osjećajem pripadnosti zajednici koja dijeli znanje i podršku.

Najveću zahvalnost izražavam svojoj obitelji. Njihova vjera u mene, strpljenje i ljubav dali su mi snagu da ustrajem, čak i u trenucima kada se činilo da su prepreke nepremostive. Vaša podrška bila je neizostavan temelj na kojem sam gradio svaki dio ovog rada.

Hvala svima od srca.

SADRŽAJ

SAŽETAK ii
ABSTRACT iii
PREDGOVOR iv
SADRŽAJ v
1. UVOD 1
1.1. Problem i predmet istraživanja 1
1.2. Svrha i ciljevi istraživanja 2
1.3. Znanstvena hipoteza i pomoćne hipoteze 3
1.4. Metodologija istraživanja 3
1.5. Očekivani znanstveni doprinos 7
1.6. Primjena rezultata istraživanja 7
1.7. Struktura doktorskog rada 8
2. PREGLED DOSADAŠNJIH ISTRAŽIVANJA 10
2.1. Procjena rizika pomorskih nezgoda: teorijski okviri i praktične metode 12
2.2. Ključni čimbenici rizika u pomorstvu 14
2.3. Tradicionalni pristupi u analizi pomorskih nezgoda 15
2.4. Primjena umjetne inteligencije u pomorstvu: trendovi i izazovi 17
2.5. Primjeri i studije slučaja primjene strojnog učenja u predviđanju nezgoda 18
2.6. Integracija senzorskih podataka i geografskih informacijskih sustava u analizi rizika 20
2.7. Potporni vektori 21
2.8. Slučajne šume 22
2.9. Neuronske mreže 23
2.10. Ansambl i slojeviti modeli 25
2.11. Sinteza dosadašnjih istraživanja i smjerovi razvoja 27
3. METODOLOŠKI OKVIR ISTRAŽIVANJA 31
3.1. Izvor, kodiranje i opis ulaznih podataka 41
3.2. Transformacija varijabli i imputacija nedostajućih vrijednosti 67
3.3. Korelacijska analiza promatranih čimbenika 70
3.4. Podjela podataka na trenirajući, validacijski i testni skup 73
3.5. Metrike procjene performansi modela i predikcije ishoda 74
3.6. Opis razvoja i treniranje modela 78
3.6.1. Opis postupka treniranja modela potpornih vektora 79

3.6.2. Opis treniranja modela slučajnih šuma i optimizacija broja stabala 80
3.6.3. Opis treniranja modela neuronske mreže i optimizacija hiperparametara..... 81
3.6.4. Opis ansambl i slojevitog modela 82
3.7. Višerazinski pristup validaciji modela 84
3.8. Izvoz rezultata i dokumentiranje procesa..... 85
4. RAZVOJ MODELA I EVALUACIJA 88
4.1. Korelacijska analiza varijabli 88
4.2. Rezultati klasifikacije, predikcije nezgoda i procjene rizika primjenom potpornih vektora..... 96
4.3. Rezultati klasifikacije, predikcije nezgoda i procjene rizika primjenom slučajnih šuma 104
4.4. Rezultati klasifikacije, predikcije nezgoda i procjena rizika primjenom neuronskih mreža..... 109
4.5. Usporedni prikaz rezultata klasifikacije, predikcije pomorskih nezgoda i procjene rizika primjenom potpornih vektora, slučajnih šuma i neuronskih mreža 118
4.6. Ansambl model 121
4.7. Slojeviti model..... 122
4.8. Integrirana usporedna analiza klasifikacijskih performansi, vjerojatnosti i procjene rizika modela 123
4.9. Provjera ponašanja modela pri kontroliranoj promjeni ciljnih oznaka 135
5. RAZMATRANJA 140
5.1. Tumačenje rezultata i usporedba modela 140
5.2. Značaj najvažnijih varijabli za pomorski rizik 142
5.3. Doprinos istraživanja 143
5.4. Ograničenja istraživanja..... 145
5.5. Preporuke i smjernice za daljnja istraživanja 146
6. ZAKLJUČAK..... 148
7. LITERATURA..... 151
POPIS KRATICA 161
POPIS SLIKA..... 165
POPIS TABLICA..... 166
PRILOZI..... 168

1. UVOD

Pomorska plovidba u obalnim vodama predstavlja složen i zahtjevan segment pomorskog prometa, jer se odvija u prostorno ograničenim područjima, pod utjecajem promjenjivih meteoroloških i oceanografskih čimbenika te u uvjetima pojačanog pomorskog prometa. Takvi uvjeti povećavaju rizik od nezgoda, koje mogu imati ozbiljne posljedice na ljudske živote, imovinu i okoliš. S obzirom na složenost i dinamičnost ovih uvjeta, razumijevanje i procjena rizika zahtijevaju integrirani pristup koji uzima u obzir tehničke karakteristike brodova, ljudski čimbenik te prirodne i operativne uvjete plovidbe. U tom kontekstu, razvoj modela koji omogućuju predviđanje i smanjenje rizika od nezgoda postaje ključan za unaprjeđenje sigurnosti i učinkovitosti plovidbe u obalnim vodama.

1.1. Problem i predmet istraživanja

Problem istraživanja bio je usmjeren na analizu, klasifikaciju i predikciju čimbenika rizika koji utječu na sigurnost plovidbe u obalnim vodama, s ciljem smanjenja broja pomorskih nezgoda. Plovidba u ovim vodama je izrazito izazovna zbog ograničenog prostora za manevriranje brodova, prisutnosti brojnih plitkih područja, blizine obale te visoke koncentracije brodova. Intenzivan pomorski promet u lukama i navigacijskim kanalima dodatno povećava vjerojatnost sudara. Nepovoljni vremenski uvjeti, poput magle, jake kiše, oluja i smanjene vidljivosti, te ljudska pogreška i tehnički kvarovi vitalnih sustava brodova, dodatno ugrožavaju sigurnost plovidbe. U sklopu istraživanja, ove prijetnje su identificirane, kvantificirane i integrirane u razvijene modele, što je omogućilo preciznu procjenu i predikciju rizika. Analizom složenih interakcija između svih čimbenika, rad pruža konkretne uvide i smjernice za unaprjeđenje sigurnosnih mjera u obalnoj plovidbi.

Tradicionalne metode procjene rizika, koje se u najvećoj mjeri oslanjaju na ručno praćenje i analizu, pokazuju ograničenu učinkovitost u pravovremenom prepoznavanju i adekvatnom reagiranju na potencijalne prijetnje. Zbog takvih metodoloških nedostataka nameće se potreba za unaprjeđenjem postojećih pristupa kroz primjenu naprednih analitičkih tehnika, među kojima posebno mjesto zauzima strojno učenje (engl. *Machine Learning* – ML). Primjena metoda strojnog učenja omogućuje znatno širi, precizniji i adaptivniji analitički kapacitet u procjeni rizika, čime se nadilaze inherentna ograničenja tradicionalnih metoda.
Osnovni problem ovog istraživanja odnosio se na potrebu sustavnog identificiranja, klasifikacije i predviđanja ključnih čimbenika rizika koji utječu na sigurnost plovidbe u obalnim vodama. Polazeći od tako definiranog problema, istraživanju se pristupilo razvojem metodološkog okvira i analitičkog sustava temeljenog na metodama strojnog učenja, čiji je cilj bio unaprijediti sigurnosne mjere te smanjiti učestalost i ozbiljnost pomorskih nezgoda. Ovakva struktura omogućuje jasan prijelaz od problema prema načinu njegova rješavanja, uz dosljednost i objektivnost u daljnjem razvoju istraživanja.

Predmet istraživanja je obuhvaćao primjenu metoda strojnog učenja za analizu, procjenu, klasifikaciju i predviđanje rizika od pomorskih nezgoda u obalnim vodama. U provedbi istraživanja metode strojnog učenja korištene su za obradu i analitičku interpretaciju podataka koji se odnose na širok skup čimbenika rizika, uključujući uvjete na moru, karakteristike brodova, navigacijske obrasce, ljudsku pogrešku, meteorološke uvjete i druge relevantne čimbenike. Primjenom algoritama strojnog učenja slučajnih šuma (engl. *Random*

Forest – RF), potpornih vektora (engl. *Support Vector Machine* – SVM) i neuronskih mreža (engl. *Neural Networks* – NNs) razvijeni su modeli koji su omogućili preciznu klasifikaciju i predviđanje potencijalnih nezgoda. Time je ostvarena pouzdana podloga za smanjenje rizika, povećanje razine sigurnosti plovidbe i unaprjeđenje procesa donošenja odluka u pomorskom sustavu.

U okviru provedenog istraživanja strojno učenje omogućilo je automatsko prepoznavanje obrazaca ponašanja brodova u različitim operativnim uvjetima, identifikaciju dominantnih čimbenika rizika te generiranje pouzdanih predviđanja korisnih za sprječavanje pomorskih nezgoda. Analiza je temeljena na opsežnim povijesnim podacima o pomorskim nezgodama, što je omogućilo razradu različitih scenarija rizika i empirijsku provjeru razvijenih modela u uvjetima koji odražavaju stvarni pomorski promet. Primjenom naprednih računalnih tehnika razvijen je sustav sposoban za brzu i preciznu klasifikaciju te procjenu rizika u skladu s trenutačnim okolnostima plovidbe.

Objekt istraživanja obuhvaćao je dva jasno definirana segmenta. Prvi segment činili su plovni objekti koji su sudjelovali u pomorskom prometu obalnih voda, uključujući komercijalne brodove, putnička i teretna plovila, manja plovila te specijalizirane brodove, pri čemu su se analizirale njihove tehničke i operativne značajke relevantne za nastanak rizika. Drugi segment obuhvaćao je obilježja operativnog okruženja obalnih voda, plovnih putova, prilaznih kanala i lučkih područja, zajedno s promjenjivim uvjetima plovidbe koji su mogli povećati izloženost riziku. Navedena područja karakteriziraju različiti uvjeti plovidbe, poput plitkih voda, uskih prolaza i gustog pomorskog prometa, što zahtijeva posebnu pažnju pri procjeni rizika. Ovakvo precizno određivanje objekata istraživanja uklonilo je neodređenosti i osiguralo jasnu analitičku osnovu za procjenu sigurnosti plovidbe.

Uz plovila i fizičke karakteristike mora, objekti istraživanja uključivali su analizu čimbenika rizika koji su mogli dovesti do nezgoda, kao što su vremenski uvjeti (magla, kiša, vjetar), ljudska pogreška (pogrešna procjena ili reakcija posade) te tehnički kvarovi vitalnih sustava brodova (kvar stroja, neispravan rad navigacijskog sustava). Osim toga, objekti istraživanja obuhvaćali su i analizu statističkih podataka o pomorskim nezgodama u obalnim vodama, s ciljem identifikacije najčešćih uzroka nezgoda i prepoznavanja kritičnih točaka u plovidbi koje su zahtijevale poboljšanje sigurnosnih mjera.

Korištenjem relevantnih podataka i naprednih modela strojnog učenja u ovom su istraživanju rizici precizno mapirani na temelju povijesnih nezgoda, stvarnih uvjeta na moru i specifičnih karakteristika brodova, što je omogućilo pouzdanu identifikaciju obrazaca koji utječu na sigurnost plovidbe. Primijenjeni pristup omogućio je analizu složenih međuodnosa između različitih čimbenika rizika te pružio jasno definirane smjernice za unaprjeđenje sigurnosnih postupaka u pomorskom sustavu. Stupanj obrade i razina pouzdanosti mapiranih rizika detaljno su prikazani u nastavku rada.

1.2. Svrha i ciljevi istraživanja

Svrha provedenog istraživanja bila je analizirati pomorske nezgode u obalnim vodama i razviti optimalne modele za procjenu rizika, s naglaskom na unaprjeđenje sigurnosti plovidbe. Obalne vode karakteriziraju uski prolazi, plitka područja, blizina obale, gust promet u lukama i kanalima te promjenjivi meteorološki i oceanološki uvjeti, što značajno povećava izloženost

riziku. Nepovoljni vremenski uvjeti, tehnički kvarovi brodova i ljudske pogreške dodatno ugrožavaju sigurnost plovidbe.

Cilj istraživanja bio je razviti znanstveno utemeljene modele strojnog učenja za precizno predviđanje rizika od pomorskih nezgoda. Ovi modeli omogućili su identifikaciju ključnih čimbenika rizika, klasifikaciju najčešćih vrsta nezgoda i formuliranje konkretnih preporuka za optimizaciju sigurnosnih mjera. Primjena razvijenih modela ukazala je na potencijal za unaprjeđenje procesa donošenja odluka na zapovjednom mostu te za povećanje ukupne razine sigurnosti pomorskog prometa u obalnim vodama, pri čemu je potrebna daljnja empirijska provjera u uvjetima operativne primjene.

Ovim pristupom istraživanje je pružilo sveobuhvatan uvid u čimbenike rizika i učinkovite mjere prevencije, stvarajući temelj za sustavno upravljanje sigurnošću u obalnoj plovidbi.

Očekuje se da će modeli omogućiti identificiranje ključnih čimbenika rizika koji najviše utječu na sigurnost plovidbe, čime će pomorski sektor moći donijeti informirane odluke u vezi s navigacijom i operacijama na moru.

1.3. Znanstvena hipoteza i pomoćne hipoteze

Glavna hipoteza (H_1):

Primjenom modela procjene rizika koji koristi metode strojnog učenja u analizi pomorskih podataka moguće je povećati točnost predviđanja rizika pomorskih nezgoda u obalnoj plovidbi u odnosu na tradicionalne metode procjene rizika.

Pomoćne hipoteze:

H_2: Najčešći uzroci pomorskih nezgoda u obalnoj plovidbi su kombinacija ljudske pogreške i nepovoljnih vremenskih uvjeta.

H_3: Integracija naprednih navigacijskih i nadzornih sustava, utemeljenih na modelima strojnog učenja, može značajno smanjiti učestalost pomorskih nezgoda kroz pravovremeno upozoravanje i podršku u odlučivanju.

1.4. Metodologija istraživanja

Prilikom istraživanja primijenjena je kombinacija kvalitativnih i kvantitativnih metoda s fokusom na analizu pomorskih nezgoda u obalnim vodama. Podaci su prikupljeni iz međunarodnih baza podataka (npr. *Global Integrated Shipping Information System* – GISIS, *European Maritime Safety Agency* – EMSA, *Marine Accident Investigation Branch* – MAIB i dr.) te su obrađeni statističkim i deskriptivnim metodama, uključujući analizu slučajeva. Istraživanje je strukturirano u tri faze: ograničenje područja istraživanja, formuliranje problema i izrada modela.

Za razvoj modela primijenjene su metode strojnog učenja, pri čemu napredni algoritmi mogu prepoznati složene obrasce u podacima i ostvariti preciznije predikcije od klasičnih statističkih metoda. Među njima, potporni vektori rade tako da pronalaze optimalnu granicu razdvajanja između klasa podataka, što omogućuje učinkovitu klasifikaciju čak i u složenim situacijama (Vapnik, 1998). Slučajne šume predstavljaju skup odluka stabala koje zajedno donose konačnu odluku, čime se smanjuje pogreška i povećava stabilnost modela (Breiman,

2001). Neuronske mreže su modelirane prema strukturi ljudskog mozga i sastavljene su od međusobno povezanih slojeva čvorova (neurona), što im omogućuje učenje i prepoznavanje vrlo složenih i nelinearnih odnosa među podacima (Goodfellow et al., 2016).

Podaci su u ovom istraživanju podijeljeni na tri skupa: trenirajući skup (70%), validacijski skup (15%) i testni skup (15%). Takva podjela omogućila je uspostavljanje ravnoteže između dostatne količine podataka za treniranje modela i zasebnih skupova podataka za validaciju i konačnu evaluaciju, čime se smanjuje rizik od prenaučenosti ili preprilagođavanja (engl. *overfitting*). Validacijski skup korišten je za fino podešavanje hiperparametara, koji se ne uče iz podataka, već unaprijed određuju strukturu modela strojnog učenja i način njegova treniranja. Za razliku od parametara modela koji se procjenjuju tijekom procesa učenja, hiperparametri se postavljaju prije treniranja te imaju ključan utjecaj na performanse i sposobnost generalizacije modela. Primjeri uključuju broj stabala u modelima slučajnih šuma, izbor funkcije jezgre i regularizacijski koeficijent u modelima potpornih vektora te broj slojeva i neurona u neuronskim mrežama. Njihova optimalna vrijednost definira se postupcima pretraživanja i validacije radi postizanja što veće prediktivne učinkovitosti modela. Konačna procjena performansi modela provedena je na testnom skupu kako bi se dobila nepristrana i pouzdana mjera točnosti i robusnosti modela. Iako bi se modeli mogli dodatno validirati i na neovisnim skupovima podataka ili korištenjem drugih pristupa, odabrana podjela i strategija evaluacije odgovaraju prirodi dostupnih podataka i ciljevima istraživanja, osiguravajući optimalnu kombinaciju točnosti procjene i praktične primjenjivosti modela (Hastie, Tibshirani & Friedman, 2009).

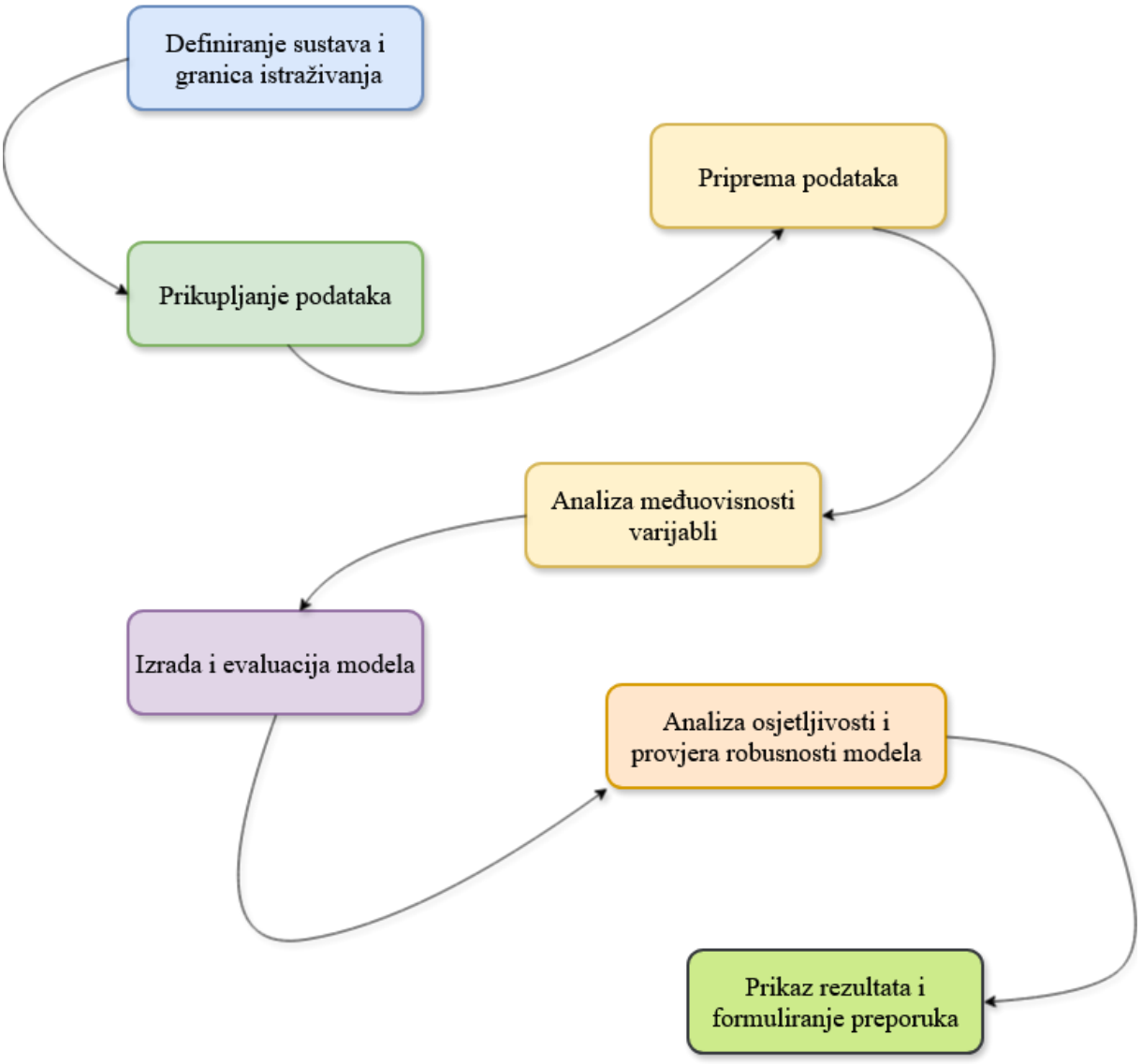
Metodološki pristup obuhvatio je čišćenje i pripremu podataka, obradu i transformaciju varijabli te analizu međuovisnosti numeričkih, tekstualnih i kombiniranih varijabli primjenom koeficijenata ovisnosti (Pearson, Spearman i Cramér) (Kuhn & Johnson, 2013; Brownlee, 2020). Nadalje, uključivao je testiranje više algoritama, verifikaciju i validaciju modela te izradu preporuka za smanjenje rizika. Ovakav pristup omogućuje analizu obrazaca u povijesnim podacima i pruža osnovu za istraživanje moguće buduće integracije u sustave potpore odlučivanju. Cilj je razvoj inteligentnih sustava za predikciju i prevenciju pomorskih nezgoda.

Metodološki pristup provedenog istraživanja obuhvaćao je nekoliko međusobno povezanih faza, koje čine cjeloviti dizajn istraživanja i omogućuju pouzdanu procjenu rizika od pomorskih nezgoda u obalnim vodama:

1. **Definiranje sustava i granica istraživanja** – jasno su definirani objekti istraživanja, uključujući plovne objekte različitih vrsta i veličina te relevantne karakteristike obalnih voda i navigacijskih puteva, s naglaskom na čimbenike koji utječu na sigurnost plovidbe.
2. **Prikupljanje podataka** – korišteni su povijesni zapisi o evidentiranim pomorskim događajima iz baza podataka opisanih u radu. Podaci nadzornih sustava u stvarnom vremenu nisu bili uključeni u empirijsku analizu.
3. **Priprema podataka** – podaci su pregledani i uređeni, uključujući provjeru nepotpunih, duplih i pogrešno unesenih zapisa. Kategorijske varijable kodirane su numerički i pretvorene u faktorske varijable, dok su ordinalne varijable pretvorene u uređene faktorske varijable. Odabrane numeričke vrijednosti razvrstane su u raspone kojima su dodijeljene brojčane oznake, a numeričke varijable standardizirane.

4. **Izrada i evaluacija modela** – razvijeni su modeli strojnog učenja, uključujući slučajne šume, potporne vektore, neuronske mreže, ansambl tj. skupne (engl. *ensemble*) i slojevite (engl. *stacking*) modele. Testirano je više algoritama kako bi se usporedile performanse u predviđanju rizika.
5. **Analiza osjetljivosti i provjera robusnosti modela** – razvijeni modeli testirani su kroz različite scenarije i uvjete plovidbe kako bi se ocijenila stabilnost i pouzdanost predviđanja, a time i primjenjivost u operativnom okruženju. Ova faza obuhvaća funkcionalnu verifikaciju modela kroz simulacije stvarnih uvjeta pomorskog prometa.
6. **Prikaz rezultata i formuliranje preporuka** – rezultati analize prikazani su u obliku grafova, tablica i dijagrama, a na temelju njih su definirane konkretne mjere za smanjenje rizika, uključujući unaprjeđenje navigacijskih protokola, edukaciju posada i integraciju naprednih tehnoloških rješenja u sustave nadzora i upravljanja plovidbom.

Shema metodološkog pristupa istraživanja prikazana je na Slici 1., koja vizualno prikazuje slijed svih faza – od definiranja sustava i prikupljanja podataka, opis podataka, opis korištenih metrika, preko pripreme i izrade modela, evaluacije i analize, do formuliranja preporuka za povećanje sigurnosti plovidbe.



Slika 1. Shema metodološkog pristupa istraživanja

Izvor: izradio autor (*diagrams.net*)

S obzirom na složenost, dinamičnost i međusobnu uvjetovanost pomorskih, meteoroloških i operativnih čimbenika sigurnosti plovidbe, u istraživanju je primijenjen skup komplementarnih znanstvenih metoda kojima je osigurana vjerodostojnost empirijskih nalaza i metodološka dosljednost. Metoda kompilacije korištena je u uvodnoj fazi istraživanja radi sustavnog pregleda postojećih znanstvenih i stručnih izvora te utvrđivanja okvira dosadašnjih spoznaja u području modeliranja rizika u obalnoj plovidbi. Metoda deskripcije omogućila je klasificiranje, opisivanje i interpretaciju ključnih činjenica, procesa i odnosa između meteoroloških, tehničkih i operativnih varijabli, čime je definiran kontekst i domena istraživačkog problema. Deduktivna metoda korištena je prilikom formuliranja istraživačkih pretpostavki zasnovanih na već potvrđenim spoznajama te u analitičkom tumačenju rezultata u odnosu prema teorijskoj osnovi.

U okviru statističke analize primijenjene su metode usmjerene na istraživanje povezanosti između varijabli, uključujući Pearsonovu korelaciju za procjenu linearne povezanosti numeričkih podataka, Spearmanovu rang-korelaciju za istraživanje monotone i neparametarske povezanosti te Cramérov V za utvrđivanje jačine veze među kategorijskim varijablama. Korelacijska analiza nije korištena za prethodnu selekciju jedinstvenog skupa ulaznih varijabli za sve modele. Pri razvoju neuronske mreže naknadno je provedena zasebna redukcija njezinih ulaza radi stabilizacije treniranja, na temelju rangiranja važnosti dobivene modelom slučajnih šuma. Postupak je opisan u dijelu koji se odnosi na neuronsku mrežu.

Primjena navedenih metoda nije bila usmjerena na smanjivanje dimenzionalnosti niti na selekciju ograničenog skupa značajki prije modeliranja. Naprotiv, u analizi su korišteni svi dostupni čimbenici, jer unaprijed nije bilo moguće sa sigurnošću utvrditi koji će model prepoznati relevantne obrasce i odnose među varijablama. Različiti modeli strojnog učenja imaju različite mehanizme učenja i osjetljivosti na ulazne podatke, pri čemu neki modeli mogu učinkovito iskoristiti linearne odnose, dok drugi bolje prepoznaju nelinearne i složene interakcije među značajkama. Stoga je zadržavanje kompletnog skupa varijabli omogućilo svakom modelu da samostalno identificira informativne čimbenike tijekom procesa učenja, bez nametanja prethodnih ograničenja koja bi mogla dovesti do gubitka potencijalno relevantnih informacija. Takav pristup osigurava veću robusnost i nepristranost u usporedbi performansi modela te povećava vjerojatnost otkrivanja skrivenih obrazaca u podacima.

Napredne metode strojnog učenja primijenjene u istraživanju obuhvaćaju potporne vektore, slučajne šume i duboke neuronske mreže, koje su omogućile prepoznavanje nelinearnih odnosa i kompleksnih interakcija unutar podataka te izgradnju prediktivnog modela rizika. Trenirajući skup korišten je za učenje modela i uspostavljanje odnosa između ulaznih parametara i pojave nezgoda, dok je validacijski skup služio optimizaciji hiperparametara pojedinih algoritama, kao što su broj stabala i maksimalna dubina kod slučajnih šuma, vrsta jezgre i regulacijski parametri kod metode potpornih vektora te broj slojeva, neurona i stope učenja kod neuronskih mreža. Testni skup korišten je isključivo za procjenu generalizacijskih svojstava modela i provjeru njegove učinkovitosti na novim, do tada neviđenim podacima, čime je osigurana nepristrana i realna procjena performansi.

Učinkovitost modela vrednovana je primjenom standardnih metrika, uključujući točnost, osjetljivost, specifičnost, F1-score i druge, što je omogućilo objektivnu usporedbu

rezultata i pouzdan odabir optimalnog modela za potporu donošenju odluka u upravljanju sigurnošću obalne plovidbe (Mishra et al., 2025). Integracija statističkih metoda za analizu povezanosti s naprednim algoritmima strojnog učenja osigurala je sveobuhvatnost analize i povećala robusnost procjene rizika u stvarnim operativnim uvjetima.

1.5. Očekivani znanstveni doprinos

U ovom istraživanju ostvaren je izravan znanstveni doprinos kroz nekoliko međusobno povezanih cjelina. Prvo, prikupljen je i obrađen opsežan skup podataka o pomorskim nezgodama u obalnim vodama, koji je potom sustavno strukturiran i standardiziran u jedinstvenu bazu podataka pogodnu za daljnju znanstvenu analizu. Time je omogućeno analitičko korištenje podataka koji prethodno nisu bili objedinjeni niti metodološki usklađeni.

Drugo, provedena je detaljna analiza prikupljenih podataka kojom su identificirani ključni čimbenici rizika te utvrđeni obrasci pojavljivanja pomorskih nezgoda. Na temelju dobivenih rezultata ostvarene su nove znanstvene spoznaje o međuodnosima varijabli koje utječu na sigurnost plovidbe u obalnim vodama.

Treće, razvijen je metodološki okvir koji uključuje primjenu metoda strojnog učenja za klasifikaciju i predviđanje pomorskih nezgoda, čime je ostvaren doprinos u području obrade složenih i heterogenih pomorskih podataka te njihove interpretacije u kontekstu procjene rizika.

Četvrto, razvijeni metodološki okvir predstavlja osnovu za buduće povezivanje modela s podacima VTS-a, Plovputa i lučkih uprava. Takvo povezivanje nije provedeno niti vrednovano u ovom istraživanju.

Slijedom navedenog, znanstveni doprinos ovog rada očituje se u uspostavi strukturirane baze podataka o pomorskim nezgodama, identifikaciji ključnih čimbenika rizika, razvoju modela temeljenih na strojnom učenju te razvoju metodološkog okvira za analizu evidentiranih pomorskih događaja, s mogućnošću njegova budućeg proširenja dodatnim institucionalnim i operativnim podacima za analizu sigurnosti pomorskog prometa.

1.6. Primjena rezultata istraživanja

Razvijeni modeli strojnog učenja predstavljaju osnovu za moguću integraciju u sustave potpore odlučivanju te omogućuju identifikaciju i kvantifikaciju rizika na temelju dostupnih podataka. Time se ne smanjuje samo ukupni rizik od nezgoda u obalnoj plovidbi, već se poboljšava planiranje navigacijskih ruta, koordinacija prometnih tokova i priprema odgovora na kritične situacije.

Potencijalni sustav potpore odlučivanju mogao bi koristiti upozorenja definirana analizom rizika u obalnoj plovidbi, primjerice predviđanje sudara, nasukavanja ili tehničkih kvarova brodova, kao i upozorenja vezana uz nepovoljne meteorološke uvjete, smanjenu vidljivost i složene oceanografske uvjete.

Buduća integracija modela s brodskim i obalnim nadzornim sustavima omogućila bi kontinuiranu razmjenu podataka o položaju broda, prometnoj situaciji i sigurnosno relevantnim parametrima unutar obalnih voda u stvarnom vremenu. U takvoj bi primjeni model mogao pružati automatske preporuke za podršku odlučivanju i upravljanje rizikom, pružajući

automatske preporuke za optimalne manevre i upravljanje rizikom, kako u rutinskim uvjetima obalne plovidbe, tako i u izvanrednim situacijama poput nezgoda, kvarova ili zagađenja mora.

Osim operativnog utjecaja, sustav ima edukativnu i stratešku vrijednost za obalnu plovidbu. Može se koristiti u simulacijama za osposobljavanje pomoraca i trening nadležnih službi, povećavajući razinu svijesti o sigurnosnim procedurama i operativnoj spremnosti. Na strateškoj razini, rezultati omogućuju razradu smjernica i propisa za integrirano upravljanje sigurnošću obalne plovidbe i optimizaciju ruta, uz smanjenje utjecaja na okoliš kroz rano otkrivanje i prevenciju potencijalnog zagađenja.

Ovakav integrirani pristup omogućio bi ne samo prevenciju nezgoda i smanjenje rizika, već i cjelokupno povećanje učinkovitosti sustava obalne plovidbe, pružajući temelje za održivi razvoj pomorstva u obalnim područjima. Takva integracija nije provedena u ovom istraživanju.

1.7. Struktura doktorskog rada

Rad je strukturiran u skladu s postavljenim istraživačkim ciljevima i metodološkim pristupom te prati logičan slijed obrade problematike povećanja sigurnosti obalne plovidbe primjenom modela procjene rizika. Istraživanje započinje uvodom, u kojem su izneseni kontekst i značaj istraživanja, identificirani glavni istraživački problemi, precizno određeni predmeti istraživanja te definirana svrha, ciljevi i istraživačke hipoteze. U ovom dijelu metodologija je dodatno razrađena, navedeni su očekivani znanstveni doprinosi te je ukazano na mogućnosti primjene rezultata u praksi. Zaključno je prikazan pregled cjelokupne strukture doktorskog rada.

Drugo poglavlje pruža pregled relevantne znanstvene i stručne literature, fokusirajući se na temeljnim teorijskim okvirima i metodama procjene rizika pomorskih nezgoda. Posebna pažnja posvećena je analizi ključnih čimbenika rizika, uključujući meteorološke uvjete, ljudski čimbenik te tehničke karakteristike brodova i pomorskih sustava, koji utječu na sigurnost plovidbe u obalnim vodama. Ovaj pregled omogućava razumijevanje postojećih pristupa procjeni rizika i identificira praznine koje opravdavaju razvoj integriranog sustava za podršku odlučivanju temeljenog na modelima strojnog učenja. Posebna pažnja je posvećena tradicionalnim pristupima analizi pomorskih nezgoda, aktualnim trendovima i izazovima u primjeni umjetne inteligencije (engl. *Artificial Intelligence* – AI) s naglaskom na strojno učenje u sigurnosnoj analizi. U poglavlju su prikazani i konkretni primjeri te studije slučaja primjene modela strojnog učenja, a posebno se ističe važnost integracije senzorskih podataka i geografskih informacijskih sustava (engl. *Geographic Information System* – GIS) u sustave za procjenu rizika.

U trećem poglavlju prikazan je sveobuhvatan metodološki okvir istraživanja, unutar kojeg je jasno razrađen cjelokupan tijek rada kroz logički definirane korake. Posebna pozornost posvećena je procesu prikupljanja, obrade i pripreme podataka, uključujući postupke čišćenja podatkovnog skupa te imputaciju nedostajućih vrijednosti primjenom najčešće frekvencije pojavljivanja za svaku pojedinu varijablu. Nakon opisa struktura i frekvencija podataka, provedena je analiza međusobnih odnosa korištenjem Pearsonovih, Spearmanovih i Cramérovih koeficijenata povezanosti. U ovom dijelu poglavlja detaljno je obrazložena i metodologija podjele podataka na skup za treniranje, validaciju i testiranje, uz navođenje razloga i metodoloških implikacija takve podjele.

U četvrtom poglavlju je prikaz razvoja modela potpornih vektora, slučajnih šuma i neuronskih mreža, zajedno s njihovom provjerom i ocjenom, primjenom četrnaest metrika kvalitete modela. Svi primijenjeni paketi u programskom okruženju R predstavljeni su u zasebnom dijelu poglavlja, uz objašnjenje njihove uloge u provedbi analitičkih i modelskih postupaka. Četvrto poglavlje posvećeno je rezultatima, uključujući procjenu vjerojatnosti pojave za svaki od sedam mogućih glavnih događaja te validaciju modela, pri čemu se prikazuju rezultati klasifikacije i predikcije rizika ostvareni primjenom pojedinih metoda. Uspoređena je njihova učinkovitost pomoću kvantitativnih pokazatelja, a provedene su i različite validacijske tehnike kako bi se potvrdila robusnost i pouzdanost dobivenih rezultata. Analiza rezultata je dala uvid u prikladnost i praktičnu primjenjivost pojedinih modela u sustavima upravljanja pomorskom sigurnošću. U završnom dijelu poglavlja provedena je usporedna evaluacija performansi pojedinačnih modela, ansambl pristupa te slojevitog meta-modela, čime je ostvareno sveobuhvatno i metodološki utemeljeno vrednovanje njihove učinkovitosti na analiziranom skupu podataka.

Peto poglavlje donosi sveobuhvatnu analizu rezultata, identifikaciju ograničenja istraživanja te daje preporuke za praktičnu implementaciju razvijenih modela. Također su definirani prijedlozi i smjernice za buduća istraživanja, s posebnim naglaskom na mogućnosti optimizacije postojećih sustava i integraciju novih tehnologija. Posebna pažnja posvećena je interpretaciji procijenjenih vjerojatnosti i pripadajućih razina rizika, s ciljem razumijevanja njihovog utjecaja na pouzdanost modela i donošenje sigurnosno utemeljenih odluka.

U zaključku su sažeti ključni nalazi istraživanja, potvrđene hipoteze i znanstveni doprinos rada. Naglašen je značaj rezultata u kontekstu unaprjeđenja sigurnosti obalne plovidbe, posebice kroz primjenu i integraciju modela strojnog učenja (potpornih vektora, slučajnih šuma i neuronskih mreža) za precizno predviđanje rizika od pomorskih nezgoda. Rad doprinosi razvoju operativnih i analitičkih metodologija za procjenu rizika u obalnoj plovidbi, uključujući sistematizaciju čimbenika rizika, klasifikaciju nezgoda i evaluaciju učinkovitosti različitih modela klasifikacije, predviđanja i procjene rizika u stvarnim uvjetima.

Na kraju rada nalazi se popis korištenih izvora, prilozi s detaljnim objašnjenjem i R kodom, popisi kratica, slika i tablica, čime se zaokružuje cjelokupna struktura doktorskog rada i omogućuje reprodukcija istraživanja.

2. PREGLED DOSADAŠNJIH ISTRAŽIVANJA

U kontekstu sigurnosti obalne plovidbe, jasno definiranje ključnih pojmova predstavlja temelj za sustavno razumijevanje i točnu interpretaciju istraživačkih rezultata. Terminološka preciznost osobito je važna pri analizi događaja koji uključuju element rizika ili ugroze sigurnosti, neovisno o tome odnose li se na pomorsku nezgodu ili nesreću. Svaki od navedenih izraza u međunarodnoj regulativi i stručnoj literaturi ima precizno određeno značenje, zbog čega njihovo razlikovanje čini nužan preduvjet za metodološki utemeljenu klasifikaciju podataka. U ovom radu izraz *pomorski događaji* koristi se kao zajednički opisni naziv za evidentirane *pomorske nesreće* i *nezgode* kada ih nije potrebno zasebno razlikovati. Izrazi *pomorska nesreć*a *i pomorska nezgoda* koriste se kada je razlika među njima važna. Izraz *pomorski incide*nt navodi se samo pri prikazu izvorne oznake engl. *marine incident* u podatkovnim bazama i ne uvodi se kao dodatna kategorija hrvatske terminologije.

Područje istraživanja u ovom radu ograničeno je na *obalnu plovidbu* (engl. *coastal navigation*), koja u općem međunarodnom smislu obuhvaća pomorsku plovidbu koja se odvija unutar unutarnjih morskih voda (engl. *internal waters*) i teritorijalnog mora (engl. *territorial sea*) obalne države, odnosno u zoni u kojoj je kopno stalno ili povremeno vidljivo, bez prelaska na otvoreno more. Iako se prostorni doseg može razlikovati među državama, u pravnom smislu ta se plovidba najčešće odvija unutar pojasa nad kojim obalna država ima potpuni ili značajan suverenitet prema Konvenciji Ujedinjenih naroda o pravu mora (UNCLOS, 1982; Republika Hrvatska, 2021). Najčešće je to plovidba brodova do 50 nautičkih milja od obale ili sigurnog utočišta, unutar koje se primjenjuju posebni propisi vezani za sigurnost, osposobljenost posade i tehničke zahtjeve broda. U pojedinim državnim jurisdikcijama pojas obalne plovidbe proteže se i izvan granica teritorijalnog mora, osobito u slučajevima kada obalna država zadržava nadležnost nad sigurnosnim aspektima u širem morskom pojasu. Takva praksa omogućuje dodatnu zaštitu pomorskog prometa u blizini obale, osobito u područjima intenzivne navigacije i složenih hidro-meteoroloških uvjeta. Obalna plovidba karakterizira se većom gustoćom pomorskog prometa, ograničenim manevarskim prostorom, prisutnošću plićina i navigacijskih ograničenja (Bi et al., 202; Liu et al., 2021).

U ovome istraživanju analizirani su isključivo sigurnosni aspekti obalne plovidbe. Otvoreno more (engl. *open sea*) nije bilo predmet analize. U skladu s Međunarodnim pravilima za izbjegavanje sudara na moru (engl. *International Regulations for Preventing Collisions at Sea* – COLREGS), predmet istraživanja su bili isključivo brodovi u plovidbi, dok su iz analize isključeni brodovi na sidru, privezani brodovi te plovila unutar brodogradilišta ili dokova jer se kao takvi ne nalaze u aktivnom režimu plovidbe (Frančić, 2012).
Osim toga, iz istraživanja su izostavljeni požari na brodu zbog njihove izrazite nepredvidivosti u kontekstu dostupnih podataka i atributa, kao i ozljede članova posade i putnika koje nisu uzročno povezane s glavnim događajima poput sudara ili nasukavanja. Takvi slučajevi, iako mogu biti ozbiljni, ne ispunjavaju uvjete za uvrštavanje u model analize rizika temeljen na navigacijskim i tehničkim čimbenicima (Chen et al., 2024).

U području sigurnosti plovidbe pojam *pomorske nezgode* definira se kao nenamjeran, iznenadan i nepoželjan događaj koji nastaje tijekom pomorskih operacija, pri čemu rezultira ili može rezultirati materijalnom štetom, narušavanjem sigurnosti ljudi i brodova, negativnim utjecajem na morski okoliš ili poremećajima gospodarskih aktivnosti povezanih s pomorskim

sektorom. Pomorska nezgoda stoga obuhvaća sve situacije tijekom plovidbe koje su dovele ili su mogle dovesti do ozbiljnih posljedica, neovisno o tome jesu li te posljedice odmah vidljive ili se manifestiraju naknadno, primjerice u vidu odgođenog onečišćenja ili naknadnih tehničkih kvarova (EMSA, 2024).

U analitičkom i stručnom kontekstu, pomorska se nezgoda strukturirano promatra kroz četiri temeljne i međusobno povezane dimenzije koje omogućuju njezinu sustavnu klasifikaciju, usporedivost i vrednovanje rizika:

1. okolnosti nastanka nezgode,
2. posljedice nezgode,
3. ozbiljnost nezgode i
4. uzroci nezgode.

Čimbenik *okolnosti* obuhvaća skup fizičkih, operativnih i kontekstualnih uvjeta pod kojima je došlo do nezgode. Tu se prvenstveno razmatraju meteorološki i oceanografski utjecaji (valovi, struje, vjetar, oborine), hidrografske karakteristike akvatorija (dubina, konfiguracija plovnih putova, podmorska infrastruktura), vidljivost i komunikacijski uvjeti. U širi operativni okvir uključuje se i gustoća pomorskog prometa, prisutnost brodova različitih karakteristika te ograničenja koja proizlaze iz specifičnosti navigacijskog područja. (EMSA, 2024).

Okolnosti se analiziraju kako bi se razumjelo u kojem je okruženju rizik eskalirao te u kojoj je mjeri vanjski čimbenik bio determinirajući element događaja, što omogućuje preciznije upravljanje rizicima, planiranje rute i optimizaciju sustava potpore odlučivanju.

Čimbenik *posljedica* odnosi se na razinu utjecaja nezgode na ljude, materijalna sredstva, okoliš i gospodarske tokove. Prema smjernicama Međunarodne pomorske organizacije (engl. *International Maritime Organisation* – IMO) i IMO metodologiji formalne procjene sigurnosti (engl. *Formal Safety Assessment* – FSA) posljedice pomorskih nezgoda procjenjuju se kroz utjecaj na ljudske živote, zdravlje, imovinu i morski okoliš (IMO, 2018).

U skladu s navedenim kriterijima, posljedice se mogu razvrstati na:

- štetu na plovilu i infrastrukturi,
- ozljede, pogibije ili evakuacija posade i putnika,
- onečišćenje mora i obalnog područja,
- ekonomske gubitke i prekid logističkih lanaca.

Cao i suradnici (2025) naglašavaju da posljedice ne smiju biti promatrane isključivo kroz trenutačni ishod, već i kroz dugoročne kumulativne efekte, poput degradacije okoliša, gubitka ekosustavnih usluga ili reputacijskih rizika za operatere i države zastave (Cao et al., 2025).

Ozbiljnost se definira prema opsegu i težini nastale štete te predstavlja ključni element u klasifikaciji pomorskih nezgoda. Ona omogućuje standardizaciju izvještavanja, statističku usporedivost i izgradnju modela procjene rizika. EMSA (2024) predviđa klasifikaciju od incidenta bez značajnijih posljedica, preko nezgoda ograničenog utjecaja, do teških i vrlo teških nezgoda s dugotrajnim ekološkim, sigurnosnim ili ekonomskim posljedicama (EMSA, 2024). Ozbiljnost čini temelj za prioritetizaciju preventivnih mjera, raspodjelu resursa i definiranje protokola reakcije u hitnim situacijama.

Čimbenik *uzroka* predstavlja najsloženiji dio analize jer uključuje interakciju ljudskih, tehničkih i okolišnih čimbenika. Prema (Onyshchenko et al., 2024), uzroci se klasificiraju u:

- tehničke, uključujući kvar pogona, navigacijskih sustava ili tehničko stanje broda,

- ljudske, povezane s procjenom, vještinom upravljanja, zamorom, komunikacijom ili nepoštivanjem procedura,
- okolišne, koje definiraju uvjeti mora, klime i okoline u trenutku operacije.

Posebno se naglašava da se uzroci rijetko svode na jedan čimbenik. Pomorske nezgode najčešće proizlaze iz kumulativnog i neočekivanog sučeljavanja više čimbenika, što zahtijeva sustavnu procjenu rizika i razvoj adaptivnih modela sigurnosti.

Promatranje pomorske nezgode kroz navedene dimenzije stvara temelj za analitički dosljedno razumijevanje rizika, omogućujući:

- prevenciju kroz predviđanje rizika,
- učenje iz povijesnih podataka,
- tehnološki i proceduralni razvoj sigurnosnih sustava,
- preciznije modeliranje u okviru prediktivne analitike i umjetne inteligencije.

Ovakav multidimenzionalni pristup predstavlja standard suvremenih sigurnosnih politika i okosnicu integriranih sustava upravljanja rizikom u maritimnom sektoru.

Međunarodna pomorska organizacija propisuje sljedeću klasifikaciju pomorskih nezgoda: one koje izazivaju vrlo teške posljedice (engl. *very serious casualty*), one s ozbiljnim posljedicama (engl. *serious casualty*), nezgode manjeg intenziteta (engl. *less serious casualty*) i pomorski incidenti (engl. *marine incidents*) koji najčešće ne rezultiraju materijalnom štetom, ali ukazuju na potencijalne sigurnosne propuste (IMO, 2010; IMO, 2003; EMSA, 2024).

U daljnjem tekstu koristi se izraz *pomorske nezgode* kao zajednički naziv za sve oblike sigurnosno-relevantnih događaja, uključujući pomorske nesreće, događaje manjeg intenziteta i manje sigurnosne propuste, sukladno međunarodnim definicijama i klasifikacijama IMO-a i EMSA-e.

Ovo razlikovanje je ključno jer omogućuje konzistentnost u izvještavanju, analizi i statističkom praćenju pomorskih događaja u okviru međunarodnog pravnog i operativnog okvira.

Sveobuhvatno definiranje korištene terminologije u ovom poglavlju čini temelj za primjenu suvremenih metoda procjene rizika u sljedećim poglavljima te osigurava metodološku dosljednost i znanstvenu utemeljenost daljnje analize.

2.1. Procjena rizika pomorskih nezgoda: teorijski okviri i praktične metode

Rizik u pomorstvu definira se kao funkcija vjerojatnoće nastanka nezgode i njenog potencijalnog utjecaja na ljudski život, plovilo i okoliš. Upravljanje rizikom stoga uključuje identifikaciju prijetnji, analizu uzročnih čimbenika i procjenu posljedica, što omogućuje donošenje informiranih odluka za prevenciju i smanjenje štetnih ishoda (Kristiansen, 2005; Knapp & van de Velden, 2024; Aven, 2012).

Dosadašnja struka u području pomorske sigurnosti razvila je teorijske okvire koji omogućuju sustavan pristup riziku. Tradicionalno, procjena rizika temelji se na statističkoj analizi povijesnih podataka o nezgodama, uključujući sudare, nasukavanja i mehanička oštećenja trupa. Takva analiza pruža empirijsku osnovu za oblikovanje preventivnih mjera, optimizaciju procedura i utvrđivanje kritičnih točaka rizika (Antão et al., 2024; Kristiansen, 2005).

Rizik u kontekstu sigurnosti obalne plovidbe predstavlja temeljni koncept bez kojeg nije moguće provesti kvalitetnu analizu, kvantifikaciju ni upravljanje sigurnosnim procesima u pomorskom okruženju. U znanstvenoj i stručnoj literaturi rizik se ne promatra kao apstraktan pojam, već kao mjerljiva veličina koja povezuje vjerojatnost nastanka neželjenog događaja i težinu njegovih posljedica. Aven navodi da je rizik znanjem uvjetovana i odluci orijentirana konstrukcija koja integrira nesigurnost i posljedice te omogućuje racionalno donošenje odluka u uvjetima neizvjesnosti (Aven, 2012). U pomorskoj domeni, taj je odnos tradicionalno definiran kao funkcija $R=f(P,C)$, pri čemu P označava vjerojatnost pojave, a C veličinu posljedica (IMO, 2010).

Ovakva definicija omogućuje primjenu kvantitativnih metoda u analizi podataka, što je osobito relevantno u području modeliranja i procjene sigurnosnih scenarija.
Suština rizika u obalnoj plovidbi proizlazi iz činjenice da se plovidbeni sustav odvija u fizički i operativno složenom okruženju koje karakteriziraju dinamični prometni tokovi, ograničeni akvatoriji, promjenjivi meteorološki uvjeti i heterogena struktura posada i brodova. Statističke analize pokazuju da se pomorske nezgode u obalnim područjima koncentriraju oko nekoliko dominantnih tipova događaja poput sudara, nasukavanja, oštećenja trupa te posljedičnih oštećenja, koji zajedno čine okosnicu ukupnog pomorskog rizika (Antão et al., 2024).

Istraživanja također potvrđuju da rizik nije isključivo funkcija tehničkih parametara, već predstavlja rezultat međudjelovanja ljudskih pogrešaka, okolišnih čimbenika i tehničkog stanja broda (Liu et al., 2021).

Formalni pristup analizi rizika u pomorstvu razrađen je kroz IMO metodu formalne procjene sigurnosti (engl. *Formal Safety Assessment* – FSA), koja zahtijeva sustavno identificiranje opasnosti, kvantificiranje rizika, procjenu prihvatljivosti te predlaganje mjera za smanjenje rizika (IMO, 1997). Formalna procjena sigurnosti ujedno naglašava da se rizik ne može promatrati izolirano, već isključivo u relaciji prema prihvatljivim kriterijima, koji odražavaju društvenu toleranciju prema mogućim posljedicama. Upravo zato pomorski sustavi koriste kontinuirano praćenje nezgoda i *izbjegnutih nezgoda* (engl. *near miss*) kao ključne indikatore rizika, jer *izbjegnuta nezgoda* reprezentira *rizik u fazi prije manifestacije* te omogućuje preventivnu reakciju (Georgoulis & Nikitakos, 2019). Premda izbjegnute nezgode predstavljaju iznimno vrijedan izvor informacija za identifikaciju sigurnosnih slabosti u sustavu, one u ovom istraživanju nisu uključene jer ne sadrže mjerljive posljedice koje bi omogućile pouzdanu kvantitativnu analizu rizika (IMO, 2008; Safety4Sea, 2022; Perkovič et al., 2024). Ipak, njihova potencijalna analitička vrijednost upućuje na mogućnost da se buduća istraživanja, potaknuta nalazima ovoga rada, usmjere i na razvoj metoda za kvantificiranje i sustavno istraživanje izbjegnutih nezgoda.

IMO u dokumentu SN.1/Circ.296 preporučuje strukturiran pristup procjeni rizika kroz faze identifikacije opasnosti, analize i evaluacije, čime se omogućuje implementacija mjera za smanjenje rizika unutar prihvatljivih granica (IMO, 2010). Paralelno, relevantna stručna i znanstvena literatura naglašava kriterije prihvatljivosti rizika u obalnim područjima, uzimajući u obzir tehničke, pravne i okolišne aspekte (Mohović, 2010).

Praktične metode procjene rizika uključuju prikupljanje i analizu podataka nacionalnih i međunarodnih agencija. Izvješća Australskog ureda za sigurnost prometa (engl. *Australian Transport Safety Bureau* −ATSB), Nizozemskog odbora za sigurnost (engl. *Dutch Safety Board*

– DSB) i Europske agencije za pomorsku sigurnost (engl. *European Maritime Safety Agency* – EMSA) služe kao empirijski temelj za analizu uzroka nezgoda i razvoj sigurnosnih modela (ATSB, 2025; DSB, 2025; EMSA, 2024).

U suvremenim pristupima procjeni rizika primjetan je trend prelaska s deskriptivnih metoda na prediktivne i dinamičke modele. Razvoj umjetne inteligencije i strojnog učenja omogućuje identificiranje ključnih čimbenika rizika i predikciju nezgoda s visokom točnošću. Metode poput potpornih vektora, slučajnih šuma i neuronskih mreža koriste se za klasifikaciju i regresiju rizika temeljenih na velikim skupovima podataka, uključujući podatke sustava za automatsku identifikaciju brodova (engl. *Automatic Identification System* – AIS), meteorološke uvjete i karakteristike broda (Knapp & van de Velden, 2024; Agorkua et al., 2024; Chen et al., 2025; Munim et al., 2024). Takvi modeli omogućuju kvantitativno vrednovanje vjerojatnoće nastanka nezgode, identifikaciju najrizičnijih ruta i optimizaciju preventivnih mjera.

Najnovija istraživanja potvrđuju da se struktura rizika u obalnoj plovidbi mijenja pod utjecajem sve veće gustoće prometa i kompleksnosti plovidbenih obrazaca, što se posebno vidi kroz analize velikih skupova podataka sustava za automatsku identifikaciju (Kim et al., 2022). Suštinski, rizik u obalnoj plovidbi treba razumjeti kao dinamičnu veličinu. On se kontinuirano mijenja u ovisnosti o taktičkim odlukama zapovjednika, trenutnim uvjetima na moru i promjenama u konfiguraciji plovnog puta. Istraživanja potvrđuju da precizna procjena rizika zahtijeva ne samo statističku deskriptivnost, nego i razumijevanje uzročnih struktura događaja, što se postiže primjenom Bayesovih mreža i topoloških analitičkih metoda (Jiang et al., 2020); (Li et al., 2023).

U konačnici, svrha analize rizika u obalnoj plovidbi nije samo kvantificirati opasnosti, već omogućiti optimizaciju odluka, smanjenje nesigurnosti i povećanje ukupne razine sigurnosti pomorskog sustava. Procjena rizika pomorskih nezgoda predstavlja temeljni element upravljanja sigurnošću plovidbe, osobito u obalnim područjima gdje su gustoća prometa i osjetljivost okoliša izrazito visoke.

Integracija tradicionalnih i suvremenih metoda procjene rizika omogućuje prelazak s reaktivnog na proaktivni pristup. Korištenje sustava za automatsku identifikaciju podataka, senzorskih mreža i geografskih informacijskih sustava omogućuje predikciju nezgoda u stvarnom vremenu, optimizaciju sigurnosnih protokola i donošenje odluka temeljenih na procjeni rizika (Durlik et al., 2024; Rawson et al., 2021).

Zaključno, procjena rizika pomorskih nezgoda razvila se u interdisciplinarni okvir koji povezuje sigurnosno upravljanje, inženjering, regulativu i tehnologiju. Sustavi procjene rizika danas omogućuju ne samo prepoznavanje postojećih prijetnji, već i anticipaciju budućih nezgoda, čime se značajno povećava sigurnost plovidbe u obalnim vodama i efikasnost preventivnih mjera.

2.2. Ključni čimbenici rizika u pomorstvu

Procjena rizika u pomorstvu zahtijeva sustavnu identifikaciju i analizu čimbenika koji značajno utječu na sigurnost plovidbe. Temeljni čimbenici rizika obuhvaćaju meteorološke uvjete, ljudski čimbenik te tehničke aspekte brodova i opreme, pri čemu svaki od njih može samostalno, ali i u kombinaciji s drugim čimbenicima, povećati vjerojatnost nastanka pomorske nezgode (Antão, 2024; Cao et al., 2025; Bye & Aalberg, 2018; Wang et al., 2021).

Meteorološki uvjeti, uključujući jak vjetar, valove, maglu i smanjenu vidljivost, predstavljaju jedan od najvažnijih izvora rizika. Nepovoljni uvjeti smanjuju preciznost navigacije i povećavaju vjerojatnost sudara ili nasukavanja, osobito u složenim plovnim područjima (Antão, 2024; Cao et al., 2025; Bye & Aalberg, 2018; Chen et al., 2019).

Ljudski čimbenik prepoznat je kao dominantan uzrok većine pomorskih nezgoda. Uključuje pogreške u odlučivanju, neadekvatnu obuku, umor, stres te nedostatak komunikacije unutar posade ili između broda i službe nadzora i upravljanja pomorskim prometom. Empirijska istraživanja pokazuju da do 70-80% pomorskih nezgoda proizlazi iz različitih oblika ljudske pogreške ili neadekvatnog upravljanja sigurnosnim procedurama (Musulin et al., 2024; Mok et al., 2023; Coraddu et al., 2020; Fan et al., 2020).

Implementacija Međunarodnog kodeksa za upravljanje sigurnošću (engl. *International Safety Management Code* – ISM Code) smanjila je utjecaj ljudskog čimbenika, ali izazovi ostaju u području kontinuirane edukacije i usvajanja novih tehnologija (Gül et al., 2020; Mok et al., 2023; Hänninen et al., 2014).

Tehnički čimbenici obuhvaćaju stanje broda, navigacijske sustave, opremu i održavanje. Kvarovi na brodu, uključujući probleme s propulzijom, upravljanjem, elektroničkom opremom i sigurnosnim sustavima, predstavljaju značajan rizik za nastanak nezgode. Redovito održavanje, modernizacija opreme i uvođenje naprednih tehnologija poput automatiziranih sustava za nadzor i predikciju kvarova ključni su za smanjenje tehničkih rizika. Statistički podaci pokazuju da se učestalost nezgoda uzrokovanih tehničkim kvarovima značajno smanjuje u sustavima sa strogim nadzorom i primjenom suvremenih metoda procjene rizika (Chen et al., 2024; Musulin et al., 2023; Wang & Yang, 2018; Li et al., 2021).

Važno je naglasiti međusobnu povezanost ovih čimbenika. Nepovoljni vremenski uvjeti mogu povećati opterećenje na ljudski čimbenik, dok tehnički kvarovi u nepovoljnim uvjetima mogu dovesti do eskalacije nezgode. Suvremeni modeli procjene rizika stoga nastoje integrirati meteorološke, ljudske i tehničke dimenzije u jedinstveni okvir, koristeći statističke metode i metode strojnog učenja za kvantifikaciju utjecaja pojedinih čimbenika (Cao et al., 2025; Liu et al., 2021; Munim et al., 2024; Rawson et al., 2021; Rawson & Brito, 2022; Atak & Arslanoglu, 2021; Park & Jeong, 2021).

Posebno su značajni pristupi temeljeni na Bayesovim mrežama koji omogućuju modeliranje nesigurnosti i međusobnih odnosa čimbenika rizika u pomorskim nezgodama (Cao et al., 2023; Wang et al., 2021; Hänninen et al., 2014; Ung, 2021; Jiang et al., 2020).

Sveobuhvatna analiza ključnih čimbenika rizika u ovom poglavlju postavlja temelj za daljnje istraživanje i razvoj preventivnih strategija. Iako su u ovom poglavlju navedeni kao temeljni element procjene rizika, detaljna analiza njihovog relativnog utjecaja, obrazaca u podacima i implikacija za operativne preporuke bit će predstavljena u poglavlju Razmatranja. Također će biti pružena detaljna interpretacija rezultata, uključujući rangiranje važnosti pojedinih čimbenika te njihova operativna značenja i implikacije.

2.3. Tradicionalni pristupi u analizi pomorskih nezgoda

Tradicionalni pristupi u analizi rizika i pomorskih nezgoda temelje se na statističkoj obradi podataka i prostorno-vremenskim metodama, koje omogućavaju sustavan uvid u uzroke i

učestalost nezgoda te predstavljaju osnovu za procjenu rizika u pomorskom prometu. Statističke metode obuhvaćaju prikupljanje, obradu i analizu podataka o nezgodama s ciljem identificiranja obrazaca, trendova i ključnih čimbenika rizika (Field, 2018).

Deskriptivne statistike omogućavaju uvid u učestalost, tipologiju i posljedice nezgoda, dok se napredni statistički modeli, uključujući regresijske i multivarijatne analize, koriste za prepoznavanje korelacija i uzročnih veza među čimbenicima (Antão, 2024; Musulin et al., 2024a; Musulin et al., 2024b; Musulin et al., 2023; Chen et al., 2025; Uyanik & Guler, 2013).

Analize koje pokrivaju razdoblje od 1990. do 2020. godine pokazuju da su sudari, nasukavanja i oštećenja trupa najčešći tipovi nezgoda, što ukazuje na potrebu usmjeravanja sigurnosnih mjera upravo na te kategorije (Antão, 2024).

Statistička obrada podataka također omogućava kvantifikaciju rizika te evaluaciju učinkovitosti postojećih sigurnosnih protokola i pravila, kao i identifikaciju sustavnih slabosti. Statistički i korelacijski pristupi, poput Pearsonovog koeficijenta, Spearmanovog rang koeficijenta i Cramérov V, koriste se za utvrđivanje povezanosti između varijabli i nezgoda, što omogućuje identifikaciju ključnih čimbenika rizika za daljnje modeliranje (Akoglu, 2018; Mukaka, 2012; Kearney, 2017).

Prostorno-vremenske metode, nadopunjujući statističke analize, koriste tehnologiju geografskih informacijskih sustava za vizualizaciju i analizu distribucije nezgoda u prostoru i vremenu. Primjena ovih metoda omogućuje identifikaciju tzv. *crnih točaka* na pomorskim plovnim putevima odnosno lokacija s povećanom učestalošću nezgoda, što je ključno za nadležne službe sigurnosti plovidbe poput službe nadzora i upravljanja pomorskim prometom i druge službe nadležne za sigurnost plovidbe. Analiza prostorno-vremenskih podataka uključuje praćenje obrazaca u odnosu na meteorološke uvjete, gustoću pomorskog prometa i druge čimbenike rizika (Yildiz et al., 2022; Musulin et al., 2023). Ove metode omogućavaju dinamičku procjenu rizika koja se može prilagoditi promjenama okolišnih i operativnih uvjeta; primjerice, integracija vremenskih uvjeta poput magle, oluja ili jakih vjetrova u modele analize pomaže u predviđanju i prevenciji nezgoda te optimizaciji plovnih ruta (Chen et al., 2024).

Iako modeli koji kombiniraju statističke i prostorno-vremenske podatke pružaju solidan temelj za procjenu rizika, oni su ograničeni analizom prošlih događaja i ne uspijevaju u potpunosti integrirati složenost suvremenih operativnih uvjeta, tehničkih inovacija i ljudskog čimbenika. Ograničenja uključuju nedostatnu dostupnost i kvalitetu podataka, te nemogućnost provođenja analize u stvarnom vremenu (Rawson et al., 2021; Rawson et al., 2022; Munim et al., 2024).

Zbog navedenih ograničenja, tradicionalni pristupi se danas nadograđuju primjenom metoda umjetne inteligencije i strojnog učenja, koje omogućuju dubinsku analizu velikih skupova podataka i razvoj dinamičnih, proaktivnih modela procjene rizika u stvarnom vremenu. Ipak, osnovna vrijednost statističkih i prostorno-vremenskih metoda ostaje u pružanju empirijski utemeljenog okvira za razumijevanje, analizu i predviđanje rizika i pomorskih nezgoda, što je i dalje ključan element za razvoj sigurnosnih standarda i pravila plovidbe (Chen et al., 2024; Munim et al., 2024).

Zaključno, statističke i prostorno-vremenske metode i dalje čine neizostavan dio sigurnosnog menadžmenta u pomorstvu, pružajući temeljne informacije o uzrocima, učestalosti i lokacijama nezgoda, te omogućuju razvoj učinkovitih preventivnih mjera i sigurnosnih

politika procjene i predviđanja rizika. Njihova integracija s modernim tehnologijama i analitičkim pristupima otvara nove mogućnosti za unaprjeđenje sigurnosti obalne plovidbe i šireg pomorskog prometa (Antão, 2024; Musulin et al., 2023; Yildiz et al., 2022).

2.4. Primjena umjetne inteligencije u pomorstvu: trendovi i izazovi

Metode strojne i duboke analize podataka omogućuju daljnje prepoznavanje čimbenika koji najsnažnije utječu na rizik, uključujući meteorološke uvjete, hidrografska ograničenja, ponašanje posade, prostornu gustoću prometa i tehničke parametre broda (Chen et al., 2025; Cao et al., 2025). Nadalje, simulacijske metode poput Monte Carlo simulacija i hidrodinamičkih modela omogućuju procjenu utjecaja različitih scenarija na sigurnost plovidbe i razinu rizika (Balas & Balas, 2025; Cao et al., 2025).

Moderni modeli strojnog učenja pokazali su visoku uspješnost u predviđanju ozbiljnosti nezgoda i dugoročnog rizika incidenata, pri čemu hibridni modeli koji integriraju hidrodinamičke podatke i Monte Carlo simulacije predstavljaju najnaprednije pristupe (Feng et al., 2024; Munim et al., 2024; Balas & Balas, 2025).

Primjena umjetne inteligencije u pomorstvu omogućava strojnim sustavima automatsko učenje iz podataka, prepoznavanje obrazaca i donošenje odluka uz minimalnu ljudsku intervenciju. U kontekstu sigurnosti plovidbe, primjena umjetne inteligencije, posebno modela strojnog učenja, predstavlja ključni iskorak u procjeni i upravljanju rizicima pomorskih nezgoda. Modeli strojnog učenja omogućavaju automatiziranu obradu velikih skupova podataka prikupljenih iz brodskih senzora, navigacijskih sustava i nadzornih platformi, što pruža uvid u složene obrasce i uzroke nezgoda. Ovi modeli nadmašuju tradicionalne statističke metode jer uče iz podataka i prilagođavaju se promjenjivim uvjetima, čime se povećava točnost procjene rizika i učinkovitost preventivnih mjera (Durlik et al., 2024; Rawson et al., 2021).

U posljednjem desetljeću metode strojnog učenja sve se češće primjenjuju u istraživanjima pomorske sigurnosti za analizu podataka o nezgodama, procjenu rizika i predviđanje mogućih događaja. Metode poput klasifikatora potpornih vektora, slučajnih šuma i neuronskih mreža se koriste za identifikaciju i predviđanje rizika i pomorskih nezgoda na temelju ulaznih podataka koji uključuju vremenske uvjete, operativni status broda, tehničke karakteristike te ljudski čimbenik (Munim et al., 2024; Feng et al., 2024; Zhang et al., 2025a; Wang et al., 2025; Zhang et al., 2025b).

Automatizirani modeli omogućavaju kontinuirano praćenje i analizu rizika u stvarnom vremenu, čime se poboljšava brzina donošenja odluka i pravovremenost reakcije na opasnosti (Zhang et al., 2025c; Liu et al., 2025; Khan et al., 2025; Portillo Juan et al., 2025).

Podaci iz sustava za automatsku identifikaciju imaju ključnu ulogu u primjeni umjetne inteligencije, jer pružaju detaljne informacije o položaju, brzini i smjeru kretanja brodova. Prostorno-vremenske analize podataka iz sustava za automatsku identifikaciju omogućuju modelima strojnog učenja detekciju ponašanja izvan uobičajenih obrazaca u plovidbi te identifikaciju situacija visokog rizika u prometno zagušenim područjima i uskim plovnim putovima (Ma et al., 2024). Integracija geografskih informacijskih sustava s tehnikama strojnog učenja dodatno omogućava precizno prostorno modeliranje rizika, što je važno za sigurnost obalne plovidbe i upravljanje pomorskim događajima u stvarnom vremenu.

Primjena strojnog učenja nosi i izazove. Prije svega, kvaliteta podataka često je ograničena zbog nepotpunosti, netočnosti ili neusklađenosti izvora, što zahtijeva složene tehnike predobrade i čišćenja podataka za osiguranje pouzdanosti modela (Durlik et al., 2024).

Također, problem interpretabilnosti modela, osobito dubokih neuronskih mreža (engl. *Deep Neural Networks* – DNNs), predstavlja izazov jer ti modeli djeluju kao *crne kutije* (engl. *black box*), što otežava razumijevanje procesa donošenja odluka i objašnjenje sigurnosnih preporuka (Durlik et al., 2024).

Dodatni izazov predstavlja potreba za prilagodljivošću modela promjenjivim uvjetima u pomorskoj plovidbi, uključujući različite vremenske prilike, gustoću prometa i promjene u tehničkom stanju brodova. Razvijaju se modeli dubokog učenja (engl. *Deep Learning* – DL) sa složenim arhitekturama, uključujući mehanizme prostorno-vremenske pažnje (engl. *Spatio – temporal attention*), koji omogućuju bolje razumijevanje dinamičnih interakcija brodova u stvarnom vremenu i preciznije predviđanje potencijalnih nezgoda (Ma et al., 2024; Feng et al., 2024). Takvi modeli integriraju višestruke izvore podataka, uključujući senzorske podatke, vremenske prognoze i povijesne informacije o nezgodama, što dodatno poboljšava točnost procjene rizika.

Integracija modela strojnog učenja sa standardima sigurnosnog upravljanja, kao što su ISM Code i Serija normi za procjenu zaštite i sigurnosti na radu (engl. *Occupational Health and Safety Assessment Series* – OHSAS), omogućava kontinuirano unaprjeđenje sigurnosnih protokola kroz analizu podataka i donošenje utemeljenih odluka (Gül et al., 2020). Takav pristup doprinosi stvaranju dinamičnih i proaktivnih sustava procjene i predviđanja rizika, čime se povećava otpornost pomorskih operacija na nepredviđene situacije.

Primjena umjetne inteligencije u pomorstvu donosi značajan iskorak u automatiziranoj analizi podataka i preciznijem upravljanju rizicima, posebno kroz modele strojnog i dubokog učenja koji nadmašuju tradicionalne metode. Unatoč brojnim prednostima, ključni izazovi uključuju ograničenu kvalitetu podataka, nedovoljnu interpretabilnost složenih modela te potrebu za njihovom stalnom prilagodbom dinamičnim uvjetima plovidbe.

2.5. Primjeri i studije slučaja primjene strojnog učenja u predviđanju nezgoda

Primjena strojnog učenja u predviđanju rizika pomorskih nezgoda sve se češće potvrđuje kroz konkretne studije slučaja koje obuhvaćaju različite algoritme, ulazne podatke i kontekste operacija. Ove primjene ilustriraju kako prediktivna sposobnost modela strojnog učenja omogućava rano upozoravanje na rizike, klasifikaciju ozbiljnosti nezgoda i donošenje odluka u stvarnom vremenu, posebno u složenim ili visoko dinamičnim pomorskim okruženjima.

U jednoj od recentnih studija primijenjen je automatizirani pristup strojnog učenja (engl. *Automated Machine Learning* – AutoML) za procjenu rizika pomorskih nezgoda. Najbolji rezultati postignuti su kombinacijom klasifikatora temeljenog na ekstremnoj gradijentnoj pojačanoj metodi (engl. *Extreme Gradient Boosting* – XGBoost) i inženjeringa značajki temeljenog na povijesnim događajima, meteorološkim uvjetima i podacima o gustoći prometa, čime je omogućeno precizno predviđanje nezgoda uz visoku razinu točnosti (Munim, 2024).

Drugi primjer uključuje primjenu poboljšanih algoritama strojnog učenja za predikciju težine nezgoda, pri čemu se uzimaju u obzir čimbenici poput vrste broda, lokacije, uzroka

nezgoda i statusa pogona u trenutku nezgode. Takvi modeli uključuju logističku regresiju, potporne vektore i slučajne šume, a rezultati su pokazali da modeli dubokog učenja nadmašuju tradicionalne metode u klasifikaciji ozbiljnosti nezgoda, osobito u složenim situacijama poput nasukavanja u uskim plovnim putovima (Feng, 2024).

U studijama provedenima u kineskim obalnim vodama analizirani su uzročni čimbenici nezgoda koristeći podatke iz stvarnih brodskih operacija, uključujući meteorološke uvjete, informacije o brodskom prometu i podatke o nezgodama. Modeli strojnog učenja omogućili su identifikaciju dominantnih uzroka nezgoda te međusobnih odnosa među ključnim čimbenicima, čime je postavljen temelj za razvoj sustava ranog upozorenja (Liu, 2021).

Napredne prostorno-vremenske metode koje kombiniraju podatke iz sustava za automatsku identifikaciju podataka s modelima usmjerene pažnje (engl. *attention mechanism*) omogućuju preciznu procjenu razmaka između brodova u prometno zagušenim područjima, što doprinosi ranom prepoznavanju situacija koje bi mogle rezultirati sudarom. Sustavi razvijeni s prostorno-vremenskom pažnjom uče iz obrazaca kretanja brodova i prepoznaju ponašanja koja odstupaju od uobičajenih obrazaca, što prethodi nezgodama (Ma, 2024).

Analiza rizika nasukavanja brodova dodatno koristi modele dubokog učenja u stvarnom vremenu, gdje kombinacija optimizacijskog postupka rojem čestica (engl. *Particle Swarm Optimization* – PSO) i dugoročno-kratkoročne memorije (engl. *Long Short-Term Memory* – LSTM) omogućava visoko preciznu procjenu rizika na temelju podataka o manevriranju, reljefu morskog dna i meteorološkim uvjetima (Zhang, 2024).

Primjena strojnog učenja na kontejnerske brodove pokazuje prednosti kombinacije statističkih modela i tehnika strojnog učenja za analizu uzroka gubitaka kontejnera. Prediktivni model obuhvaća varijable poput parametara stabilnosti broda, hidrometeoroloških uvjeta i tehničkih karakteristika tereta, a rezultati ukazuju na značajno poboljšanje u predviđanju rizičnih situacija, što omogućava optimizaciju sigurnosnih procedura pri ukrcaju i plovidbi (Yi, 2025).

Također, modeli slučajnih šuma primijenjeni su u netradicionalnim scenarijima sigurnosti, uključujući nepredviđene manevre, utjecaj ljudskog čimbenika u blizini luka i složene hidrometeorološke promjene. Uspješnost tih modela potvrđena je u simulacijama i retrospektivnim analizama nezgoda u područjima pomorskog prometa velike gustoće (Lu, 2022).

Analiza korelacije pomorskih nezgoda i sigurnosnih aktivnosti u Jadranu pokazuje da integracija statističkih metoda i algoritama strojnog učenja omogućava razumijevanje odnosa između kontrole plovidbe, edukacije posade i učestalosti nezgoda. Ovim se potvrđuje da strojno učenje nije samo alat za predikciju, već i sredstvo za evaluaciju učinkovitosti postojećih mjera pomorske sigurnosti (Musulin, 2024).

Sve ove studije slučaja potvrđuju sposobnost strojnog učenja modela u prepoznavanju skrivenih obrazaca, integraciji različitih izvora podataka i proaktivnom upravljanju rizikom. Njihova primjena pruža temelje za razvoj integriranih sigurnosnih sustava koji omogućavaju pravovremene informacije donositeljima odluka i operativnom osoblju u sektoru pomorstva, čime se značajno povećava sigurnost plovidbe.

2.6. Integracija senzorskih podataka i geografskih informacijskih sustava u analizi rizika

U suvremenoj analizi rizika u obalnoj plovidbi sve značajniju ulogu imaju integracija senzorskih podataka i geografskih informacijskih sustava, čime se omogućuje prostorno-vremensko modeliranje i sustavna procjena pomorskih nezgoda. Dosadašnja postignuća u struci pokazuju da kombinacija preciznih senzorskih informacija i geolokacijskih podataka omogućuje identificiranje ključnih čimbenika rizika te razumijevanje obrazaca nastanka nezgoda u stvarnom operativnom kontekstu.

Podaci iz sustava za automatsku identifikaciju brodova, radarskih mreža, meteoroloških senzora i dubinomjera pružaju višedimenzionalne informacije o kretanju brodova, stanju mora i okolišnim uvjetima, što je omogućilo detaljnu analizu rizičnih situacija u obalnoj plovidbi (Zhang, 2024). Dosadašnja praksa pokazuje da je vizualizacija zona povećanog rizika i identificiranje uskih grla u prometnim tokovima ključna za prevenciju nezgoda i planiranje sigurnosnih mjera, posebice u područjima plitkih prolaza ili područja povećane gustoće prometa.

Integracija podataka s geografskim informacijskim sustavima omogućuje identifikaciju uzročnih veza između tehničkih i ljudskih čimbenika te okolišnih uvjeta, stvarajući osnovu za dinamičko modeliranje rizika. Primjena dubokih neuronskih mreža s prostorno-vremenskom pažnjom omogućava predviđanje vremenskih razmaka između brodova (engl. *time headways*) u stvarnom vremenu, čime se omogućuje rana identifikacija rizičnih susreta brodova i mogućnosti sudara (Ma, 2024).

Kombinacija podataka iz sustava za automatsku identifikaciju i batimetrijskih informacija, primijenjena u prediktivnim modelima za rizik nasukavanja, potvrđuje važnost povezivanja povijesnih obrazaca kretanja brodova s karakteristikama manevriranja i okolišnim uvjetima (Zhang, 2022). Integracija meteoroloških i radarskih podataka dodatno omogućuje identifikaciju kritičnih čimbenika rizika u uskim pomorskim prolazima, gdje geolokacijska analiza pomaže prepoznati sigurnosne prijetnje povezane s vidljivošću, strujama i prometnom gustoćom (Yildiz, 2022).

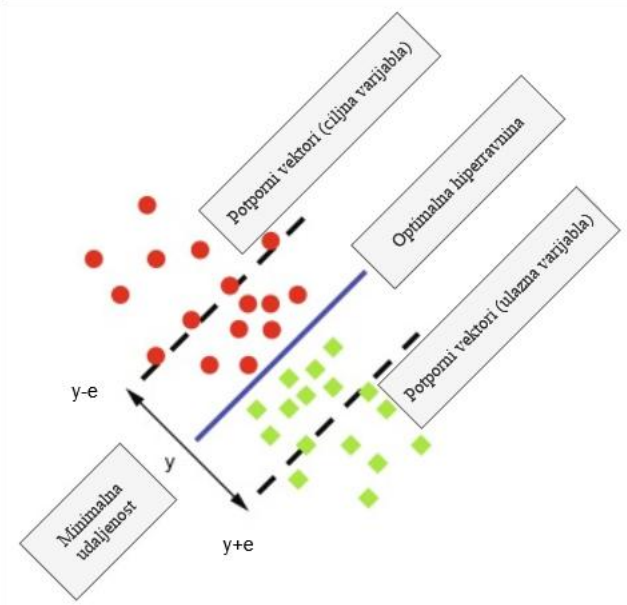
Dosadašnja postignuća u primjeni višerazinskih modela učenja s pojačanjem (engl. *Reinforcement Learning* – RL) pokazuju kako se integracija više senzorskih izvora može koristiti za autonomno planiranje plovidbenih ruta i procjenu optimalnih putanja u stvarnom vremenu, uzimajući u obzir moguće pomorske nezgode i sigurnosne čimbenike (Waltz, 2025).

Ovi pristupi nadograđuju prethodno prikazanu statističku i prostorno-vremensku analizu i predviđanje rizika nezgoda, čime se omogućuje ne samo retrospektivno razumijevanje obrazaca, već i proaktivno upravljanje rizikom. Integracija senzorskih sustava i geografskih informacijskih sustava stoga predstavlja temelj za razvoj inteligentnih sustava prediktivnog i preventivnog djelovanja u obalnoj plovidbi, doprinosi boljoj procjeni rizika i omogućuje donošenje učinkovitih operativnih i strateških odluka (Rawson, 2022).

Na ovaj način, poglavlje povezuje dosadašnja postignuća u struci s primjenom suvremenih tehnologija, pokazujući kako se teorijski prikazi rizika i nezgoda sada mogu praktično primijeniti u kontekstu prediktivne analize i proaktivnog upravljanja sigurnošću plovidbe.

2.7. Potporni vektori

Metoda potpornih vektora predstavlja jednu od najvažnijih i najraširenijih tehnika strojnog učenja, koja se koristi za klasifikaciju i regresiju u različitim primjenama, uključujući procjenu rizika (Vapnik, 1998). Temeljna ideja metode je pronalaženje optimalne hiperravnine koja razdvaja podatke različitih klasa s najvećom mogućom marginom. Hiperravnina je geometrijska površina u višedimenzionalnom prostoru koja prostor dijeli na dva dijela, dok optimalna hiperravnina predstavlja onu površinu koja najbolje razdvaja podatke različitih klasa, uz najveću udaljenost od najbližih podataka, odnosno potpornih vektora (Slika 2).



Slika 2. Grafički prikaz hiperravnine i potpornih vektora u SVM klasifikaciji
Izvor: (Roy & Kasiviswanathan, 2023)

Margina između klasa predstavlja udaljenost između hiperravnine i najbližih podataka različitih klasa. Maksimalizacija margine povećava pouzdanost klasifikacije novih podataka i smanjuje mogućnost pogrešnog razvrstavanja (Lin et al., 2025).

Za nelinearno odvojive probleme metoda koristi funkcije jezgri, poput linearne, polinomne ili radijalne bazne funkcije, koje omogućuju preslikavanje podataka u prostor više dimenzionalnosti, gdje je razdvajanje klasa izvedivo (Vapnik, 1998).

U primjeni na pomorsku sigurnost potporni vektori koriste se za identifikaciju i klasifikaciju rizičnih situacija na temelju složenih skupova podataka, uključujući položaj broda, brzinu, smjer, vremenske uvjete i gustoću prometa (Chen et al., 2024; Ma et al., 2024).

Prednost modela potpornih vektora ogleda se u sposobnosti rada s visoko dimenzionalnim i nepotpunim podacima te u otpornosti na prenaučenost zahvaljujući maksimalizaciji margine (Agorkua et al., 2024). Za višeklasne probleme primjenjuju se strategije jedan-naspram-ostalih i jedan-naspram-jednoga (Hsu & Lin, 2002; Rifkin & Klautau, 2004).

Kombinacija s drugim tehnikama strojnog učenja omogućuje precizniju klasifikaciju događaja i prioritizaciju situacija visokog rizika (Munim et al., 2024; Liu et al., 2021; Yi et al., 2025).

Model potpornih vektora pruža visoku točnost, robusnost i jasnu interpretaciju klasifikacijskih granica, što omogućuje pravovremenu aktivaciju sigurnosnih protokola i optimizaciju operacija u stvarnom vremenu (Lin et al., 2025).

Formalno, funkcija odlučivanja metode potpornih vektora (Cortes & Vapnik, 1995) definirana je kao:

$$f(x) = w^T x + b \tag{1}$$

pri čemu se predikcija klase određuje kao:

$$\hat{y} = sign(f(x)) \tag{2}$$

U slučaju nelinearno odvojivih podataka, funkcija odlučivanja proširuje se korištenjem funkcije jezgre (engl. *kernel function*):

$$f(x) = \sum_{i=1}^{n} \alpha_i \, y_i K(x_i, x) + b \tag{3}$$

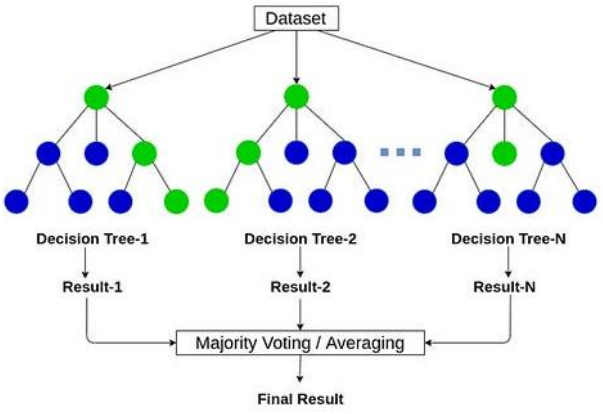
gdje je $K(x_i, x)$ funkcija jezgre, a α_i Lagrangeovi multiplikatori. Za višeklasnu klasifikaciju primjenjuje se strategija jedan-naspram-ostalih, pri čemu se za svaku klasu $c \in \mathcal{C}$ trenira zaseban klasifikator. U ovom istraživanju klasifikacija se provodi za sedam klasa pomorskih nezgoda. Sukladno tome, skup klasa označava se kao $\mathcal{C} = \{1, \dots, 7\}$. Konačna predikcija definira se kao:

$$\hat{y} = arg\, max_{c \in \mathcal{C}} f_c(x), \ \mathcal{C} = \{1, \dots, 7\} \tag{4}$$

Metoda potpornih vektora stoga se pokazuje iznimno učinkovitom u klasifikaciji rizičnih situacija u pomorskom okruženju, jer omogućuje razdvajanje složenih, višedimenzionalnih podataka uz visoku točnost i robusnost. Njezina primjenjivost u stvarnim uvjetima potvrđuje važnost ove metode kao pouzdanog alata za pravovremeno donošenje sigurnosnih odluka i optimizaciju operativnih procesa.

2.8. Slučajne šume

Slučajne šume su metoda strojnog učenja koja kombinira više stabala odlučivanja kako bi se poboljšala točnost i robusnost modela (Liaw & Wiener, 2002). Svako stablo gradi se na nasumično odabranim podskupovima podataka i značajki, a konačna odluka određuje se konsenzusom između svih stabala kao što je prikazano na Slici 3. Ovaj pristup smanjuje prenaučenost i povećava sposobnost modela za generalizaciju na nove podatke.



Slika 3. Općenita struktura modela slučajnih šuma
Izvor: (Omarzai, 2024)

Formalno, u kontekstu višeklasne klasifikacije, gdje je skup mogućih klasa definiran kao $\mathcal{C} = \{1,2,3,4,5,6,7\}$, slučajne šume procjenjuju vjerojatnost pripadnosti svakoj klasi kao udio stabala koji predviđaju određenu klasu:

$$P(Y = c \mid x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(T_b(x) = c), c \in \{1, \dots, 7\} \quad (5)$$

gdje je $T_b(x)$ predikcija b-tog stabla, B ukupan broj stabala u modelu, a $\mathbb{I}(\cdot)$ indikator funkcija koja poprima vrijednost 1 ako je uvjet ispunjen, odnosno 0 ukoliko nije. Konačna predikcija definira se kao klasa s najvećom procijenjenom vjerojatnošću (Breiman, 2001):

$$\hat{y} = arg\,max_{c \in \mathcal{C}} P(c \mid x) \quad (6)$$

Algoritam slučajnih šuma koristi se za predviđanje složenih ishoda i identifikaciju ključnih čimbenika rizika u različitim problemima klasifikacije i regresije (Yang, 2023). Metoda je posebno učinkovita u radu s heterogenim, nepotpunim i neuređenim podacima te može otkriti uzročno-posljedične odnose među varijablama bez potrebe za prethodnom pretpostavkom o obliku distribucije podataka (Du, 2024; Musulin et al., 2023).

Glavna prednost slučajnih šuma je u stabilnosti i robusnosti modela, kao i u mogućnosti interpretacije važnosti pojedinih značajki. Takav pristup omogućuje identifikaciju čimbenika koji imaju najveći utjecaj na predviđanje ishoda i pruža potporu donošenju odluka u operativnom planiranju i prevenciji rizika (Feng et al., 2024; Yang, 2023).

U kontekstu pomorske sigurnosti, slučajne šume primjenjuju se za analizu trajektorija brodova, vremenskih uvjeta, stanja mora i ljudskog čimbenika. Modeli omogućavaju identifikaciju rizičnih situacija, predviđanje ozbiljnosti nezgoda i podršku u donošenju preventivnih mjera (Feng et al., 2024; Lu et al., 2022).

Slučajne šume predstavljaju pouzdanu i robusnu metodu za analizu složenih i heterogenih podataka, omogućujući precizno prepoznavanje ključnih čimbenika rizika i predviđanje sigurnosno značajnih ishoda. U području pomorske sigurnosti ova metoda osigurava vjerodostojnu identifikaciju rizičnih situacija i pruža snažnu analitičku podršku u donošenju preventivnih i operativnih odluka.

2.9. Neuronske mreže

Neuronske mreže predstavljaju metodu umjetne inteligencije inspiriranu funkcijom i strukturom ljudskog mozga. Sastoje se od međusobno povezanih čvorova, nazvanih neuroni, raspoređenih u slojeve: ulazni sloj prima podatke, skriveni slojevi obrađuju informacije, a izlazni sloj daje rezultat modela (Goodfellow et al., 2016). U ovom istraživanju neuronske mreže korištene su kao jedan od modela strojnog učenja za klasifikaciju pomorskih nezgoda. Svaka veza između neurona ima pridruženu težinu koja se prilagođava tijekom procesa učenja, čime mreža uči prepoznavati obrasce i relacije u podacima.

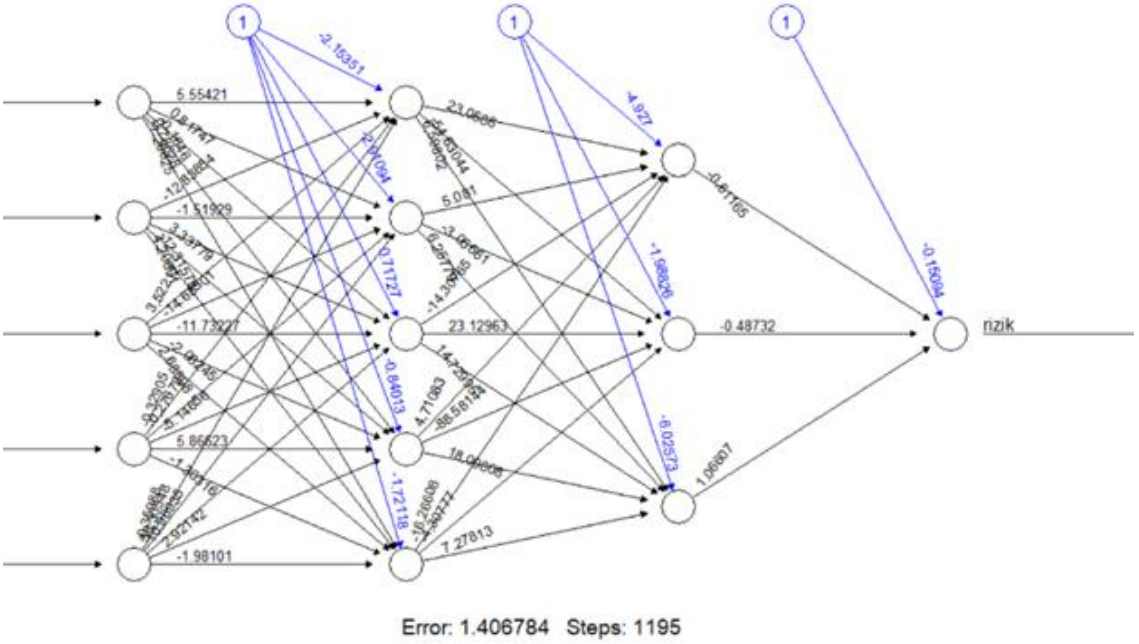
Matematički, izlaz pojedinog neurona može se prikazati kao:

$$y = f(\sum_{i=1}^{n} w_i\, x_i + b) \quad (7)$$

gdje su x_i ulazne varijable, w_i težine veza, b bias (pomak), a f nelinearna aktivacijska funkcija.

Na slici 4. prikazan je primjer arhitekture višeslojne neuronske mreže (ulazni, skriveni i izlazni sloj) s označenim težinama veza i vrijednostima aktivacija tijekom procesa učenja.

Čvorovi predstavljaju neurone, a veze među čvorovima ponderirane prijenose signala. Neuronske mreže koriste funkcije aktivacije koje omogućavaju modeliranje nelinearnih odnosa između ulaznih i izlaznih varijabli.



Slika 4. Grafički prikaz arhitekture višeslojne neuronske mreže

Izradio: autor

Najčešće korištene funkcije uključuju sigmoidnu, rektificiranu linearnu jedinicu i tangens hiperboličku funkciju. Sigmoidna funkcija definira se kao:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{8}$$

Sigmoidna funkcija pretvara ulaznu vrijednost u izlaz između 0 i 1, korisno za modeliranje vjerojatnosti i binarnih odluka, dok rektificirana linearna jedinica (engl. *Rectified Linear Unit* – ReLU) definirana je kao:

$$f(x) = max(0, x) \tag{9}$$

te daje izlaz jednak ulaznoj vrijednosti ako je veći od nule, a nulu inače, što ubrzava učenje i smanjuje problem zasićenja gradijenta. Tangens hiperbolička funkcija (*tanh*) daje izlaz između –1 i 1, omogućavajući centriranje podataka oko nule i poboljšavajući stabilnost modela u skrivenim slojevima. Proces učenja temelji se na optimizaciji težina kroz algoritme poput gradijentnog spuštanja, uz minimizaciju funkcije pogreške koja se često definira kao srednja kvadratna pogreška (Back, 2018; Garson, 1991):

$$J = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

gdje je y_i stvarna vrijednost, a $\hat{y}_i$ predviđena vrijednost modela. Ažuriranje težina tijekom učenja provodi se prema izrazu:

$$w = w - \eta\frac{\partial J}{\partial w} \tag{11}$$

gdje η označava stopu učenja (engl. *learning rate*).

Glavna prednost neuronskih mreža je njihova sposobnost učenja složenih nelinearnih odnosa u visokodimenzionalnim podacima te dobra generalizacija na nepoznate podatke. U kontekstu ovog istraživanja to omogućuje prepoznavanje kompleksnih obrazaca u podacima o pomorskim nezgodama, gdje interakcije između meteoroloških, operativnih i navigacijskih čimbenika nisu linearne. Mreže su primjenjive u klasifikaciji, regresiji i predviđanju sekvenci podataka, što ih čini posebno pogodnim za predviđanje rizika i klasifikaciju rizičnih situacija u pomorskoj sigurnosti (Goodfellow et al., 2016).

U analitičkom pristupu, neuronske mreže omogućavaju sustavno modeliranje podataka prije njihove primjene u specifičnim domenama. Tek nakon što su modeli naučeni i evaluirani, moguće je primijeniti ih u konkretnim problemima, poput predviđanja rizika u pomorskoj sigurnosti ili klasifikacije rizičnih situacija. Njihova sposobnost generalizacije i prilagodbe novim podacima pruža snažnu podršku u donošenju informiranih i pravovremenih odluka (Feng et al., 2024; Du, 2024).

2.10. Ansambl i slojeviti modeli

Kod klasifikacije rizika pomorskih nezgoda, uz pojedinačne modele poput modela potpornih vektora, slučajnih šuma i neuronskih mreža, često se primjenjuju ansambl modeli koji kombiniraju predikcije više algoritama strojnog učenja. Takav pristup omogućuje iskorištavanje prednosti pojedinačnih modela, smanjuje rizik od preučenja te povećava točnost, robusnost i pouzdanost klasifikacije pomorskih događaja u obalnoj plovidbi (Brownlee, 2021; scikit-learn, n.d.).

Glavni cilj ovakvog pristupa je povećanje generalizacijske sposobnosti modela kroz objedinjavanje različitih metodoloških perspektiva. U ovom istraživanju kombiniraju se modeli potpornih vektora, slučajnih šuma i neuronskih mreža, čime se postiže bolja pokrivenost složenih i nelinearnih obrazaca u podacima te pouzdanija procjena rizika pomorskih nezgoda (Goodfellow et al., 2016; Raschka, 2018; Lan et al., 2024).

Ansambl modeli kombiniraju predikcije više pojedinačnih modela kako bi se poboljšala ukupna točnost i smanjila varijabilnost predikcija. Općenito, ansambl model može se prikazati kao:

$$\hat{y} = \sum_{m=1}^{M} w_m \, f_m(x) \tag{12}$$

gdje je $f_m(x)$ izlaz pojedinog modela, w_m njegova težina, a M ukupan broj modela u ansamblu.

Najčešće korištene metode uključuju *bagging* (engl. *bootstrap aggregating*), *boosting* i većinsko glasanje. *Bagging* smanjuje varijancu modela generiranjem više podskupova podataka i treniranjem modela na svakom od njih, dok *boosting* iterativno trenira modele tako da se svaka sljedeća instanca usmjerava na pogreške prethodnih, čime se poboljšava ukupna točnost modela. U slučaju *boostinga*, model se može prikazati kao:

$$F(x) = \sum_{m=1}^{M} \alpha_m \, f_m(x) \tag{13}$$

gdje α_m predstavlja težinu pojedinog modela u ansamblu.

Kod većinskog glasanja, predikcije pojedinačnih modela kombiniraju se tako da klasa koju odabere najviše modela postaje konačna odluka. Matematički, ovaj pristup može se zapisati kao:

$$\hat{y} = \text{mode}\{f_1(x), f_2(x), \dots, f_M(x)\} \quad (14)$$

gdje funkcija $mode$ označava najčešće odabranu klasu među modelima.

U implementaciji većinsko glasanje se realizira kombiniranjem predikcija potpornih vektora, slučajnih šuma i neuronskih mreža putem skupne predikcije, pri čemu se za svaki zapis bira klasa koju je najviše modela označilo, a zatim se kreira matrica zabune za ansambl model. Ovaj pristup je jednostavan, ali učinkovit kada su pojedinačni modeli komplementarni, jer svaka metoda u ansamblu može ispraviti slabosti drugih modela (Hastie et al., 2009).

Slojeviti modeli predstavljaju napredniji pristup ansambliranju, u kojem se predikcije više baznih modela koriste kao ulaz za meta-model koji uči kako ih optimalno kombinirati. Za razliku od jednostavnog glasanja, meta-model prepoznaje obrasce točnosti i pogrešaka pojedinih modela te na temelju toga donosi konačnu odluku (Wolpert, 1992; Raschka, 2018).

U ovom istraživanju slojeviti pristup temelji se na kombinaciji modela potpornih vektora, slučajnih šuma i neuronskih mreža, pri čemu se njihove predikcije koriste za treniranje meta-modela. Time se omogućuje bolje iskorištavanje raznolikosti modela, smanjenje pristranosti i varijance te povećanje stabilnosti i točnosti predikcije na neviđenim podacima (Goodfellow et al., 2016; Brownlee, 2021; scikit-learn, n.d.).

Ovakav pristup posebno je učinkovit u složenim višeklasnim problemima, gdje pojedinačni modeli ostvaruju slične performanse, ali različite obrasce pogrešaka, pa njihova kombinacija rezultira pouzdanijom i robusnijom klasifikacijom. Slojeviti modeli koriste meta-model koji uči iz predikcija više baznih modela. U ovom pristupu, bazni modeli prvo generiraju predikcije na trening skupu, koje se zatim koriste kao ulazne varijable za meta-model. Matematički, slojeviti model može se prikazati kao:

$$\hat{y} = g(f_1(x), f_2(x), \dots, f_M(x)) \quad (15)$$

gdje funkcija g predstavlja meta-model koji kombinira izlaze baznih modela. Meta-model je često jednostavniji model poput logističke regresije ili stabla odluke. Primjena slojevitih pristupa omogućava modelu da uči iz sistematskih pogrešaka pojedinačnih algoritama i poboljša ukupnu točnost, posebno u problemima s kompleksnim interakcijama između ulaznih varijabli. Slojeviti modeli se mogu implementirati korištenjem posebnih funkcija, gdje predikcije baznih modela služe kao novi skup podataka za meta-model. Ovakav pristup omogućava procjenu granice razdvajanja koji svaki bazni model daje konačnoj predikciji, čime se može kvantitativno procijeniti doprinos različitih modela u ansamblu (Wolpert, 1992).

Primjena ansambl modela u klasifikaciji rizika pomorskih nezgoda omogućava preciznije predikcije i detaljniju analizu važnosti ulaznih varijabli, što je ključno za identifikaciju čimbenika rizika i planiranje preventivnih mjera u obalnoj plovidbi. Osim kreiranja ansamblnih predikcija, izvodi se i analiza važnosti varijabli za pojedinačne modele, što doprinosi boljem razumijevanju koji čimbenici imaju najveći utjecaj na procjenu rizika. Kombinacija više modela ne samo da poboljšava točnost, već i povećava robusnost sustava, smanjujući vjerojatnost pogrešnih klasifikacija važnih događaja, što je ključno za donošenje odluka u sigurnosnim i preventivnim aktivnostima u pomorstvu (Dietterich, 2000).

2.11. Sinteza dosadašnjih istraživanja i smjerovi razvoja

Dosadašnja istraživanja u području pomorske sigurnosti i procjene rizika fokusirana su na identifikaciju, kvantifikaciju i predviđanje rizičnih situacija kako bi se smanjile posljedice nezgoda. Rizik se definira kao kombinacija vjerojatnoće nastanka nezgode i ozbiljnosti njenih posljedica, dok se nezgoda smatra neželjenim događajem koji može prouzrokovati štetu ljudima, plovilima, okolišu ili ekonomskim interesima (Vapnik, 1998; Lin et al., 2025).

Okvirna istraživanja pokazala su da je analitički pristup riziku ključan za pravovremeno donošenje odluka. Tradicionalne metode uključuju statističku obradu povijesnih podataka, analizu trendova nezgoda, te identifikaciju čimbenika koji doprinose pomorskim događajima. Međutim, s razvojem velikih i heterogenih podataka, klasične metode pokazale su ograničenja u prepoznavanju složenih obrazaca i predviđanju rizičnih situacija (Chen et al., 2024; Ma et al., 2024).

U novijim istraživanjima sve veći značaj dobivaju metode strojnog učenja, koje omogućavaju klasifikaciju i predviđanje rizičnih događaja na temelju višedimenzionalnih podataka uključujući položaj broda, brzinu, smjer, vremenske uvjete, gustoću prometa i ljudski čimbenik. Metode poput potpornih vektora, slučajnih šuma i neuronskih mreža omogućavaju preciznu identifikaciju i klasifikaciju rizičnih situacija, dok istovremeno pružaju robusnost modela i mogućnost generalizacije na nove, nepoznate uvjete (Agorkua et al., 2024; Feng et al., 2024; Goodfellow et al., 2016).

Ovi pristupi potvrđuju da kombinacija teorijskih okvira rizika i modernih metoda strojnog učenja omogućava pravovremeno prepoznavanje opasnosti, optimizaciju operativnih procesa i podršku u donošenju preventivnih i sigurnosnih mjera u pomorskoj praksi.

Tablica 1. daje strukturirani pregled istraživanja o procjeni rizika u obalnoj plovidbi analiziranih u ovom radu. Prikazani su glavni teorijski pristupi i modeli procjene rizika, kao i praktične metode koje se primjenjuju u upravljanju sigurnošću plovidbe, ključni čimbenici rizika, glavni doprinosi i ograničenja pojedinih istraživanja te autori i godine objave. Posebno su istaknuti: tradicionalni statistički pristupi temeljeni na analizi nezgoda, suvremeni modeli koji koriste umjetnu inteligenciju i AIS podatke, formalna procjena sigurnosti prema IMO-u, uloga ljudskog, tehničkog i okolišnog čimbenika te koncept *izbjegnutih nezgoda*. Tablica sažeto pokazuje razvoj od deskriptivnih analiza prema dinamičkim, prediktivnim i proaktivnim pristupima sigurnosti obalne plovidbe.

Tablica 1. Ključna istraživanja o procjeni rizika u obalnoj plovidbi analizirana u ovom radu

Fokus istraživanja	Metode / pristupi	Ključni čimbenici rizika	Glavni doprinosi istraživanja	Ograničenja / napomene	Autor(i), godina
Definicija i teorijski okvir pomorskog rizika	Pregled literature, konceptualna analiza	Rizik, nezgode, izbjegnute nezgode	Jasno definiranje pojmova, okvir za daljnje istraživanje	Ograničena primjena u analitičkim modelima	Vapnik (1998); Lin et al. (2025)
Temeljni koncepti pomorskog rizika i sigurnosti	Statistička analiza nezgoda, teorijski modeli	Vjerojatnost događaja i posljedice	Definicija rizika, utemeljenje analitičkog pristupa sigurnosti	Osnovan na povijesnim podacima	Kristiansen (2005)
Teorijski okvir rizika kao znanjem uvjetovane konstrukcije	Konceptualni model rizika	Nesigurnost, posljedice, odluke	Razumijevanje rizika u uvjetima neizvjesnosti	Apstraktan, manje primijenjen na obalnu plovidbu	Aven (2012)
Struktura pomorskih nezgoda u obalnoj plovidbi	Statistička analiza tipova nezgoda	Sudari, nasukavanja, oštećenja trupa	Empirijska slika dominantnih rizika	Oslonac na povijesne zapise	Antão et al. (2024)
Identifikacija čimbenika rizika u obalnoj plovidbi	Statistička analiza, eksperimentalna evaluacija	Gustoća prometa, meteorološki uvjeti, ljudski čimbenik, stanje broda	Prepoznavanje ključnih varijabli koje utječu na rizik	Nepotpuni ili heterogeni podaci	Chen et al. (2024); Ma et al. (2024)
Klasifikacija nezgoda	Analiza povijesnih podataka, statističke metode	Tip nezgode, ozbiljnost posljedica	Kategorizacija nezgoda prema ozbiljnosti i tipu	Klasične metode ne prepoznaju nelinearne obrasce	Feng (2024); Lu (2022)
Procjena vjerojatnosti i težine rizika	Regresijske analize, probabilistički modeli	Vjerojatnost nastanka nezgode, težina posljedica	Kvantitativna procjena rizika	Ograničena preciznost u dinamičnim uvjetima	Yi (2025); Jiang et al. (2020)
Ljudski čimbenik u pomorskim nezgodama	Empirijske analize, modeli ponašanja	Umor, stres, obuka, komunikacija	Potvrda dominacije ljudske pogreške	Teško mjerljiv psihološki aspekt	Musulin et al. (2024); Mok et al. (2023); Coraddu et al. (2020)
Integracija senzorskih i GIS podataka	GIS, integrirani sustavi, digitalni blizanci,	Položaj broda, gustoća prometa, vremenski uvjeti	Dinamična i stvarnovremenska procjena rizika	Nedostatak standardiziranih podataka i interoperabilnosti	Ma (2024); Yuan (2024); Thakur et al. (2025)
Uzročni odnosi u pomorskim nezgodama	Bayesove mreže, topološke metode	Uzročno-posljedični lanci	Dubinska analiza mehanizama nezgoda	Složeni modeli	Jiang et al. (2020); Li et al. (2023)

Primjena umjetne inteligencije u procjeni rizika	Strojno učenje, duboko učenje, hibridni modeli	Višedimenzionalni podaci, ljudski čimbenik	Precizna klasifikacija, predikcija i generalizacija	Potreba za standardiziranim skupovima podataka	Liu (2021); Rawson (2022)
Primjena AI u procjeni rizika	Strojno učenje, analiza AIS podataka	Plovidbeni obrasci, gustoća prometa	Prediktivni modeli rizika	Potreba za velikim skupovima podataka	Knapp & van de Velden (2024)
Mogućnosti daljnjeg istraživanja i unaprijeđenja	Analiza literature, integracija modela strojnog učenja	Promjenjivi okolišni i operativni čimbenici	Identifikacija smjerova daljnjih istraživanja	Ograničena operativna primjenjivost	Feng (2024); Zhang (2024)
Utjecaj tehničkih i okolišnih čimbenika	Kvantitativno modeliranje, strojno učenje	Meteorologija, tehnički kvarovi	Predikcija nezgoda, dijagnostika kvarova	Ovisnost o podacima senzora	Cao et al. (2025); Wang et al. (2025); Chen et al. (2025)
Potporni vektori	SVM, funkcije jezgre, višeklasne strategije	Položaj broda, brzina, smjer, meteorološki uvjeti	Precizna klasifikacija rizičnih situacija, otpornost na prenaučenost	Ograničenja kod vrlo nelinearnih i složenih obrazaca	Vapnik (1998); Lin et al. (2025); Chen et al. (2024)
Slučajne šume	RF, ansambl stabala odlučivanja	Heterogeni podaci, trajektorije brodova, ljudski čimbenik	Robusna identifikacija ključnih čimbenika rizika	Mogućnost smanjenja interpretabilnosti	Liaw & Wiener (2002); Feng et al. (2024); Yang (2023)
Neuronske mreže	*Feedforward*, slojeviti perceptroni, LSTM	Veliki i višedimenzionalni podaci, nelinearni obrasci	Prepoznavanje složenih obrazaca, generalizacija na nove podatke	Složenost treniranja i potreba za velikim skupovima podataka	Goodfellow et al. (2016); Back (2018); Feng et al. (2024)
Ansambl i slojeviti modeli	Ansambl modeli (*bagging, boosting, majority voting*), *stacking*	Komplementarni modeli, interakcije varijabli	Poboljšana ukupna točnost, analiza važnosti varijabli	Složenija implementacija, zahtjev za velikim podacima	Hastie et al. (2009); Wolpert (1992); Dietterich (2000)

Izvor: izradio autor

Dosadašnja istraživanja jasno pokazuju da je procjena rizika pomorskih nezgoda temeljni element upravljanja sigurnošću obalne plovidbe. Rizik se više ne promatra samo kao statična vjerojatnost događaja, već kao dinamična veličina koja se mijenja pod utjecajem ljudskog čimbenika, tehničkog stanja brodova, gustoće prometa i okolišnih uvjeta. Razvoj metoda procjene rizika prošao je put od jednostavne statističke analize nezgoda do naprednih modela koji uključuju umjetnu inteligenciju, Bayesove mreže i analizu velikih podataka iz AIS sustava. Formalna procjena sigurnosti prema IMO-u stvorila je okvir za sustavno prepoznavanje i smanjenje rizika, dok suvremeni tehnološki razvoj omogućuje prelazak s reaktivnog na proaktivni pristup, od analize posljedica prema predikciji i prevenciji nezgoda. Posebno se potvrđuje ključna uloga ljudskog čimbenika te potreba za stalnom edukacijom i modernizacijom tehničkih sustava.

Ovo poglavlje pokazuje da je sigurnost obalne plovidbe interdisciplinarno područje koje povezuje tehnologiju, regulativu, upravljanje i pomorsku praksu. Integracija tradicionalnih i novih metoda procjene rizika stvara temelj za izgradnju učinkovitih modela koji mogu doprinijeti povećanju sigurnosti obalne plovidbe, što je i osnovni cilj daljnjih istraživanja.

3. METODOLOŠKI OKVIR ISTRAŽIVANJA

Metodološki okvir istraživanja koncipiran je s ciljem osiguravanja potpune očuvanosti informacija, znanstvene transparentnosti i metodološke konzistentnosti u procjeni vjerojatnosti pojave pomorskih nezgoda te posljedično kvantifikacije rizika u obalnoj plovidbi. Polazište metodologije temeljilo se na inherentnoj složenosti pomorskih nezgoda, koje nastaju kao rezultat djelovanja višestrukih, međusobno povezanih čimbenika, što zahtijeva sustavan, višefazni i analitički utemeljen pristup obradi podataka. U tom kontekstu, zadržavanje svih inicijalnih varijabli predstavljalo je ključni metodološki princip, budući da eliminacija varijabli u ranim fazama može dovesti do gubitka latentnih, ali operativno značajnih informacija koje modeli mogu identificirati tek kroz kompleksne međusobne odnose.

Cjelokupni metodološki postupak prikazan je u Tablici 2. i Slici 5., pri čemu je svaka faza jasno strukturirana kako bi se osigurao logičan slijed analitičkih koraka i potpuna transparentnost istraživačkog procesa.

Početne faze (1–3) obuhvaćale su unos, validaciju i pripremu podataka o nezgodama u obalnoj plovidbi, uključujući definiranje ciljne varijable *glavni događaj*. Proces čišćenja, standardizacije i transformacije podataka omogućio je homogenizaciju skupa podataka i njegovu pripremu za daljnje statističke analize i modeliranje.

Tablica 2. Pregled metodološkog okvira analize podataka istraživanja

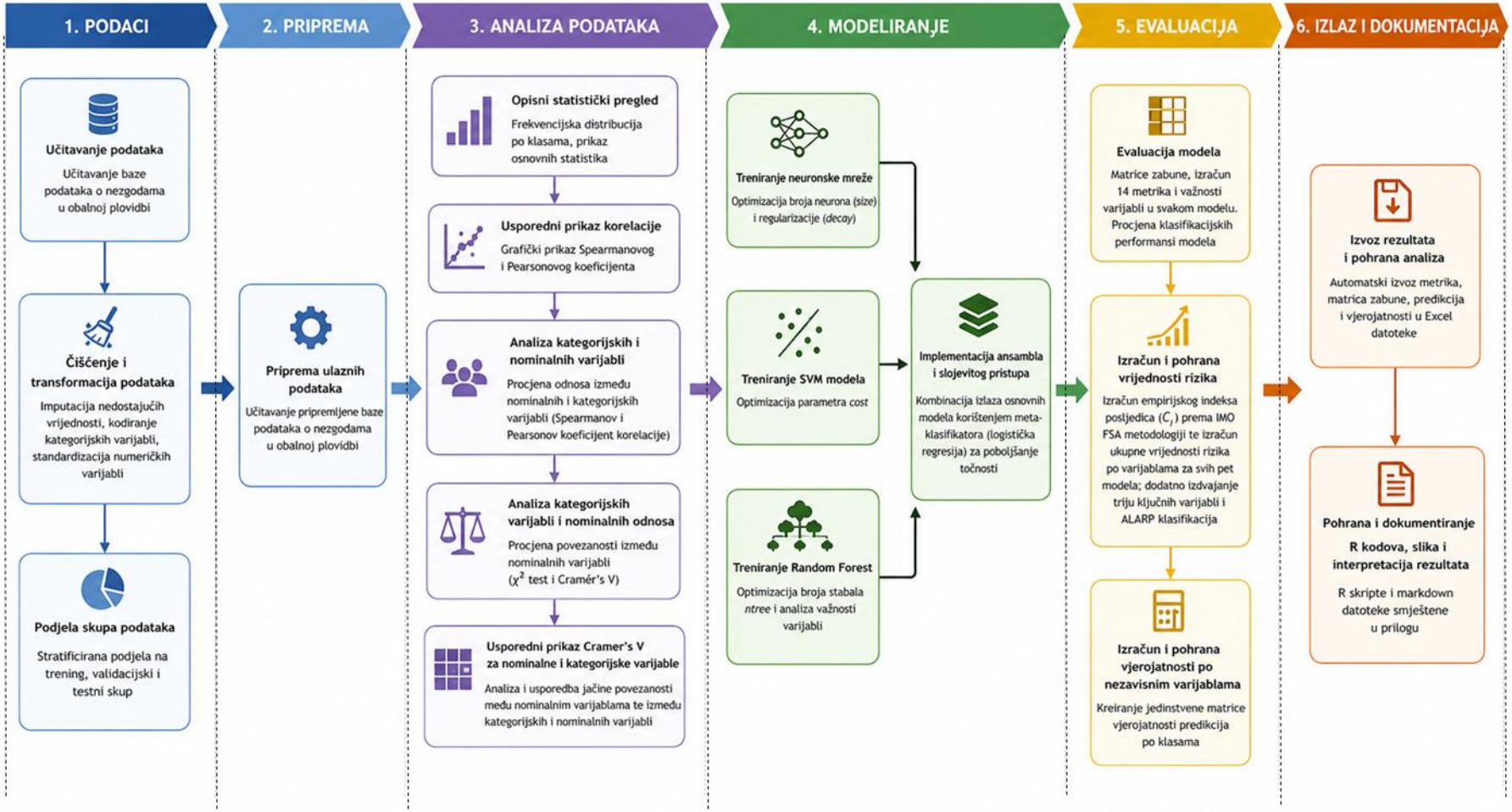
Faza	Opis aktivnosti	Tehnike i alati u R	Namjena i mjesto u radu	Komentar
1. Izvor, opis i priprema ulaznih podataka	Učitavanje baze podataka o nezgodama u obalnoj plovidbi	*readxl, dplyr, tidyr, tibble*	Početak metodološkog poglavlja; definicija ciljne varijable *main_event*	Ova faza predstavlja temeljni korak u kojem se svi relevantni podaci iz izvornog izvora unose u programsko okruženje. Korištenje pouzdanih paketa omogućava učinkovitu manipulaciju podacima, čime se osigurava njihova spremnost za daljnju obradu. Definiranje ciljne varijable je ključno jer ona predstavlja ishodišnu vrijednost koju modeli trebaju predvidjeti.
2. Čišćenje i transformacija podataka	Transformacija nedostajućih vrijednosti, kodiranje kategorijskih varijabli, standardizacija numeričkih varijabli	*janitor, dplyr, scale*	Unifikacija skupa podataka za strojno učenje	Čišćenje i transformacija podataka sprječavaju negativan utjecaj nepravilnosti na performanse modela. Imputacija nedostajućih vrijednosti i standardizacija numeričkih varijabli smanjuju pristranosti i osiguravaju dosljedan format podataka. Pravilno kodiranje kategorijskih varijabli sprječava pogrešne interpretacije modela.
3. Opisni statistički pregled	Frekvencijska distribucija po klasama, prikaz osnovnih statistika	*summary, table, ggplot2*	Prikaz osnovnih karakteristika skupa podataka	Ovaj korak pruža uvid u osnovne značajke skupa podataka, kao što su raspodjela klasa i statistički sažeci. Tako se može prepoznati neravnoteža podataka,

				potencijalni outlieri i druge karakteristike važne za planiranje daljnjih analiza.
4. Korelacijska analiza faktorskih i numeričkih varijabli	Procjena odnosa između faktorskih (kategorijskih) i numeričkih varijabli pomoću Spearmanovog i Pearsonovog koeficijenta korelacije	Spearmanov i Pearsonov koeficijent korelacije – *cor, ggcorrplot, corrplot*	Detekcija potencijalne multikolinearnost, redundantnosti i nelinearnih odnosa	Ova faza identificira međusobne povezanosti varijabli. Spearmanov koeficijent korelacije omogućuje prepoznavanje monotone, a Pearsonov koeficijent linearne povezanosti između varijabli. Time se mogu prepoznati redundantne varijable koje mogu utjecati na rezultate modela, što predstavlja osnovu za daljnju selekciju značajki.
5. Usporedni prikaz Spearmanove i Pearsonove korelacije	Usporedna analiza linearnih (Pearson) i monotonih (Spearman) povezanosti među numeričkim i ordinalnim varijablama, grafička prezentacija rezultata	*cor.test, ggcorrplot, corrplot*	Identifikacija različitih vrsta povezanosti za precizniju selekciju prediktora	Usporedni prikaz omogućava bolju interpretaciju odnosa među varijablama, otkriva skrivene nelinearne obrasce i potvrđuje koje varijable vrijedi zadržati u modelu, povećavajući njegovu točnost i stabilnost.
6. Analiza kategorijskih varijabli i nominalnih odnosa	Procjena povezanosti između nominalnih varijabli) pomoću χ^2 testa i Cramérov V koeficijenta	*χ^2 test (chisq.test), Cramérov V (assocstats* iz *vcd* paketa)	Identifikacija značajnih asocijacija i kvantifikacija jačine povezanosti između nominalnih i kategorijskih varijabli	Ovaj korak omogućava statističku potvrdu povezanosti između kategorijskih podataka. Cramérov V dodatno kvantificira snagu veze, što pomaže u razumijevanju strukture podataka i optimizaciji izbora varijabli za modele.
7. Usporedni prikaz Cramérov V za nominalne i kategorijske varijable	Detaljna analiza i usporedba jačine povezanosti među nominalnim varijablama te između kategorijskih i nominalnih varijabli	*Cramérov V,* kontingencijske tablice, grafički prikazi	Optimizacija selekcije ulaznih značajki i detekcija redundantnih varijabli	Omogućava precizno razlikovanje varijabli koje imaju visoku ili nisku međusobnu povezanost, što direktno utječe na kvalitetu modela i sprječava nepotrebno povećanje dimenzionalnosti.
8. Podjela skupa podataka	Stratificirana podjela skupa u trening (70% – 1115), validacijski (15% – 240) i testni skup (15% – 236)	*caret::createDataPartition*	Osiguranje nepristrane evaluacije modela i održavanje proporcionalne zastupljenosti	Podjela osigurava da modeli budu trenirani i evaluirani na odvojenim podacima, čime se sprječava prenaučenost. Stratifikacija održava balans klasa u svim skupovima, što je važno za stabilnost i pravednost evaluacije.
9. Treniranje modela SVM sa 7 različitih vrijednosti hiperparametra	Treniranje modela potpornih vektora nad trening skupom, optimizacija parametra *cost*	*caret, kernlab*	Komparativna analiza performansi modela u poglavlju o modeliranju	Metoda potpornih vektora pruža snažnu klasifikaciju putem pronalaska optimalne hiperravnine. Precizno podešavanje hiperparametra *cost* balansira između prenaučenosti i generalizacije, što je ključno za kvalitetne predikcije.

10. Treniranje modela RF sa 5 različitih vrijednosti hiperparametra	Izgradnja ansambl modela temeljenog na stablu odlučivanja, optimizacija broja stabala *ntree* (120, 300, 360, 480 i 500)	*randomForest, caret*	Usporedba s ostalim metodama klasifikacije i analiza važnosti varijabli	Slučajne šume smanjuju varijancu agregiranjem rezultata više stabala. Pet vrijednosti parametra *ntree* odabrano je kao praktičan opseg pretrage za usporedbu točnosti pri različitom broju stabala. Najbolja među ispitanim vrijednostima odabrana je prema najvećoj točnosti na validacijskom skupu, dok analiza važnosti varijabli pomaže u razumijevanju modela.
11. Treniranje modela NNs sa 2 različita hiperparametra ukupno 12 različitih kombinacija	Treniranje višeslojne perceptronske mreže (3 skrivena sloja), optimizacija broja neurona (*size*) i regularizacije (*decay*)	*nnet, caret*	Uključivanje nelinearnih uzoraka u predikciji rizika i analiza performansi	Neuronske mreže mogu modelirati složene nelinearne odnose, no zahtijevaju preciznu optimizaciju hiperparametara da bi se izbjegla prenaučenost ili problemi u treniranju. Postupak smanjenja značajki omogućava uspješno treniranje i u zahtjevnijim uvjetima. Za optimizaciju odabrana su dva hiperparametra jer određuju kapacitet mreže i stupanj regularizacije. Broj neurona u skrivenom sloju (*size*) određuje sposobnost mreže za modeliranje složenih odnosa, dok parametar *decay* penalizira velike težine i time pomaže ograničiti prenaučenost. Njihovim zajedničkim podešavanjem tražena je ravnoteža između prilagodbe podacima i generalizacije.
12. Evaluacija modela putem matrice zabune i izračun 14 metrika i važnosti varijabli u svakom modelu	Procjena klasifikacijskih performansi modela (*Accuracy, Precision, Recall, F1, Specificity, NPV, FPR, FNR* i dr.)	*confusionMatrix* iz caret paketa	Detaljna višedimenzionalna procjena kvalitete klasifikacije u 7 klasa	Korištenje velikog broja metrika omogućava sveobuhvatnu procjenu modela, naročito u višeklasnom problemu, gdje sama točnost nije dovoljna za razumijevanje performansi.
13. Izračun i pohrana vjerojatnosti po nezavisnim varijablama	Kreiranje jedinstvene matrice vjerojatnosti predikcija po klasama za svaku nezavisnu varijablu iz svih modela	(SVM, RF, NN) *svm(..., probability=TRUE), predict(..., type=prob), nnet(..., type=raw), openxlsx, dplyr*	Analiza detaljnih vjerojatnosti predikcije i važnosti varijabli	Ova faza omogućava detaljnu procjenu performansi modela na razini pojedinačnih varijabli i klasnih predikcija, čime se poboljšava interpretacija modela i razumijevanje njihovog ponašanja na testnim podacima.
14. Izračun i pohrana	Izračun empirijskog indeksa posljedica (C_i) te izračun ukupne	*openxlsx, dplyr, tidyr, ggplot2,*	Završni analitički korak u	Ova faza predstavlja središnji korak procjene rizika jer povezuje procijenjene vjerojatnosti pojave pojedinog događaja s

vrijednosti rizika	vrijednosti rizika po vrijednostima i po varijablama za svih pet modela; dodatno izdvajanje triju ključnih varijabli	*group_by, summarise, pivot_longer, merge, cut*	procjeni rizika; povezivanje procijenjenih vjerojatnosti s težinom posljedica i kvantifikacij a rizika po varijablama i modelima	empirijski određenim indeksom posljedica. Time se dobiva kvantitativna vrijednost rizika za svaku kategoriju i varijablu, omogućuje usporedba među modelima. Posebna obrada ključnih varijabli dodatno olakšava interpretaciju najutjecajnijih čimbenika u obalnoj plovidbi.
15. Grafička interpretacija rezultata	Vizualizacija korelacije, važnosti varijabli i dijagrama performansi modela	*ggplot2, plotly*	Vizualna podrška i interpretacija kvantitativni h nalaza	Grafički prikazi pomažu u intuitivnom razumijevanju modela i podataka, olakšavaju komunikaciju rezultata te otkrivaju obrasce koji možda nisu vidljivi u numeričkim podacima.
16. Izvoz rezultata i pohrana analiza	Automatski izvoz metrika, matrica zabune, predikcija i vjerojatnosti u Excel datoteke	*writexl, openxlsx*	Transparentn ost, reproducibil nost i dokumentaci ja cijelog postupka	Dokumentacija u standardiziranom formatu omogućava naknadnu provjeru, analizu i ponavljanje istraživanja, što je ključno za znanstvenu validnost.
17. Implementacij a *ansambl* i *slojevitog* pristupa i usporedni prikaz svih 5 modela	Kombinacija izlaza osnovnih modela primjenom meta-modela slučajnih šuma u slojevitom pristupu te usporedba s ansambl pristupom.	*caretEnsembl, caretStack*	Poboljšanje stabilnosti i generalizacij e modela strojnog učenja	Ansambliranje modela povećava robusnost i točnost predikcija jer spaja snage pojedinačnih modela, smanjujući rizik od prevelike ovisnosti o jednoj metodi.
18. Pohrana i dokumentiranj e korištenih R kodova, slika te interpretacija rezultata	Sve R skripte i markdown datoteke smješteni u prilogu	*.R i .Rmd* datoteke	Osiguranje znanstvene i tehničke reproducibil nosti istraživanja	Čuvanje koda osigurava potpunu transparentnost metodologije i olakšava ponavljanje ili nadogradnju istraživanja u budućnosti.

Izvor: izradio autor



Slika 5. Grafički prikaz metodološkog okvira istraživanja

Izvor: izradio autor (*diagrams.net*)

U fazama 4–7 provedena je detaljna analiza međusobnih odnosa varijabli primjenom korelacijskih i asocijacijskih metoda. Pearsonov i Spearmanov koeficijent korišteni su za identifikaciju linearnih i monotonih odnosa među numeričkim i ordinalnim varijablama, dok su hi-kvadrat test i Cramérov V omogućili kvantifikaciju povezanosti među kategorijskim varijablama. Ovakav pristup omogućio je dubinsko razumijevanje strukture podataka bez smanjenja broja varijabli, čime su očuvana njihova interpretacijska i prediktivna svojstva.

U fazi 8 provodena je stratificirana podjela podataka na trening, validacijski i testni skup, čime je osigurana nepristrana procjena sposobnosti generalizacije modela uz očuvanje distribucije klasa.

U fazama 9–11 implementirani su modeli strojnog učenja, uključujući metode potpornih vektora, slučajnih šuma i neuronskih mreža, uz sustavnu optimizaciju hiperparametara s ciljem postizanja optimalnog kompromisa između točnosti i generalizacijske sposobnosti.

Evaluacija modela (faza 12) provodena je primjenom matrice zabune i skupa komplementarnih metrika, čime je omogućena višedimenzionalna procjena performansi u kontekstu višeklasne klasifikacije.

U fazi 13 provodena je kvantifikacija vjerojatnosti pojave pojedinog glavnog događaja za svaku nezavisnu varijablu i njezine vrijednosti. Nakon klasifikacije nezgoda u sedam klasa glavnih događaja, za svaki primjer iz testnog skupa određena je raspodjela vjerojatnosti pripadnosti pojedinoj klasi. Kod metode potpornih vektora, udaljenosti od razdvajajuće hiperravnine transformiraju se u vjerojatnosti primjenom sigmoidne funkcije uz pristup jedan-protiv-svih (engl. *one-vs-rest*) za višeklasni problem. Kod slučajnih šuma vjerojatnosti se određuju relativnim udjelom stabala koja predviđaju pojedinu klasu, dok neuronske mreže generiraju vjerojatnosti primjenom *softmax* funkcije na izlaznom sloju. Time je dobivena konzistentna i usporediva procjena vjerojatnosti pojave za sve modele. Dobivene vjerojatnosti pojave zatim su se agregirale po nezavisnim varijablama, pri čemu se za svaku varijablu i njezine kategorije izračunavala prosječna vjerojatnost pojave svakog glavnog događaja. Ovakva agregacija omogućila je identifikaciju varijabli i njihovih vrijednosti koje su bile povezane s većim ili manjim vjerojatnostima pojave nezgoda, čime se dobio detaljan uvid u strukturu utjecaja pojedinih čimbenika.

U fazi 14 provedena je kvantifikacija rizika, pri čemu se rizik definirao kao očekivana vrijednost posljedica, odnosno: gdje je P_i modelom procijenjena vjerojatnost pripadnosti promatranog zapisa (i)-toj klasi glavnog događaja, uz uvjet da zapis pripada skupu evidentiranih pomorskih događaja. Stoga P_i ne predstavlja bezuvjetnu vjerojatnost nastanka pomorskog događaja tijekom pojedine plovidbe, na određenoj ruti ili u određenom vremenskom razdoblju. Povezivanjem procijenjenih vjerojatnosti pripadnosti klasama s empirijski definiranim indeksima posljedica C_i (engl. *severity index*) dobiva se izvedeni pokazatelj (R), koji omogućuje usporedbu relativne razine posljedica među promatranim vrstama događaja. U kontekstu procjene rizika u pomorstvu, često se koristi

matrica rizika (Tablica 3.) kao praktičan i vizualno razumljiv alat za rangiranje i ocjenu rizika pomorskih nezgoda. Kao i kod svake procjene rizika, ovakav pristup temelji se na tri osnovna koraka:

1. identifikacija opasnosti,
2. procjena učestalosti pojave neželjenih događaja,
3. procjena posljedica tih događaja.

Crveno - Neprihvatljivo (engl. *Intolerable*)
Žuto – Podnošljivo (engl. *Tolerable*)
Zeleno - Neznatno (engl. *Negligible*)

Tablica 3. Matrica rizika

	Posljedice			
Učestalost	Male	Značajne	Teške	Katastrofalne
Vrlo učestalo				VISOKI RIZIK
Vjerojatno				
Malo vjerojatno				
Izuzetno malo vjerojatno				
Gotovo nemoguće	MALI RIZIK			

Izvor: izradio autor prema (IMO, 2018; ProtectUK, 2025)

Na temelju empirijske distribucije posljedica dobiven je indeks očekivane posljedice C_i za svaki od 7 glavnih događaja definiranih u Tablici 7. (sudar, udar, nasukanje itd.). Razine učestalosti F određuju se diskretizacijom procijenjene vjerojatnosti pojave događaja P u pet kategorija (F1–F5). Kriteriji razvrstavanja indeksa posljedica i učestalosti prema FSA metodologiji prikazani su u Tablici 4.

Tablica 4. Kriteriji klasifikacije razine učestalosti i posljedica prema FSA metodologiji

Parametar	**Vrijednost (raspon)**	**Razina**	**Opis**
C_i	> 3	C4	Katastrofalne posljedice
	1 – 3	C3	Teške posljedice
	0,1 – 1	C2	Značajne posljedice
	< 0,1	C1	Male posljedice
F	> 0,2	F5	Vrlo često
	0,1 – 0,2	F4	Često
	0,05 – 0,1	F3	Umjereno često
	0,01 – 0,05	F2	Rijetko
	< 0,01	F1	Vrlo rijetko

Napomena: Vrijednosti prikazane u Tablici 4. predstavljaju kontinuirane raspone koji se koriste za klasifikaciju u diskretne razine učestalosti i posljedica, dok se cjelobrojne vrijednosti rizika dobivaju tek u fazi formiranja matrice rizika.

U okviru matrica rizika, razine vjerojatnosti (učestalosti pojave događaja) obično se određuju na temelju podataka o učestalosti pomorskih nezgoda iz statističkih analiza i kvantificiraju se kroz unaprijed definirane kategorije, što omogućuje sustavno određivanje gdje se svaki događaj svrstava unutar matrice rizika. Takva klasifikacija vjerojatnosti i procjena posljedica omogućuje daljnju procjenu rizika pomoću matrice rizika te kategorizaciju rizika kao:

- mali rizik,
- srednji rizik,
- veliki rizik.

Prema standardnoj praksi:

- mali rizici se obično mogu prihvatiti bez dodatnih mjera ili samo uz osnovno praćenje,
- srednji rizici zahtijevaju mjere za njihovo smanjenje na prihvatljivu razinu,
- veliki rizici smatraju se neprihvatljivima i zahtijevaju značajno smanjenje do razine koja je prihvatljiva i s kojom su svi sudionici svjesno suglasni.

Ovakav pristup strukturiranog vrednovanja rizika podržan je i smjernicama Međunarodne pomorske organizacije za FSA (IMO, 2018).

Ovaj dokument opisuje FSA kao sistematsku i strukturiranu metodologiju procjene rizika i evaluacije mjera kontrole rizika, koja uključuje identifikaciju opasnosti, analizu rizika, identificiranje mogućnosti kontrole rizika i procjenu ekonomske isplativosti za donošenje odluka u okviru IMO regulatornog procesa. U FSA metodologiji, učestalost i posljedice događaja definiraju se i kvantificiraju na jasan, dosljedan način, čime se olakšava objektivna procjena rizika i usporedba alternativa za kontrolu rizika. Pored kvalitativne procjene rizika, moguće je provesti i kvantitativnu procjenu rizika (engl. *Quantitative Risk Assessment* – QRA), koja predstavlja napredni pristup u kojem se rizik izražava numerički primjenom matematičkih modela, statističkih metoda i analize podataka. Takav pristup je u skladu s metodologijom FSA razvijenom od strane IMO-a, koja u okviru analize rizika uključuje primjenu kvantitativnih tehnika za procjenu frekvencije i posljedica pomorskih nezgoda. QRA omogućuje pretvaranje neizvjesnosti i povijesnih podataka o nezgodama u kvantificirane vrijednosti koje su korisne za planiranje sigurnosnih mjera i donošenje strateških odluka, a često se koristi u složenim analizama gdje je potrebno detaljnije razumijevanje opterećenja rizika i učinkovitosti mjera kontrole (IMO, 2018).

Za razliku od klasičnog FSA pristupa u kojem se učestalost događaja temelji isključivo na empirijskoj frekvenciji, u ovom radu vjerojatnosti pojavljivanja događaja procijenjene su primjenom modela strojnog učenja. Na taj način omogućena je točnija i

robusnija procjena učestalosti, temeljena na višedimenzionalnim podacima i utjecajnim varijablama sustava. Prediktivni model korišten je za:

- identifikaciju značajnih varijabli rizika,
- izračun individualnih doprinosa,
- validaciju empirijske distribucije rizika.

$$R_i = P_i \times C_i \tag{16}$$

gdje je:

- R_i – doprinos pojedine varijable ukupnom riziku analizirane pomorske nezgode,
- P_i – vjerojatnost pojave događaja,
- C_i – indeks ozbiljnosti posljedice.

Ukupni rizik za određenu pomorsku nezgodu:

$$R_k = \sum R_i \tag{17}$$

gdje je k = glavni događaj.

Empirijska procjena posljedica C_i (engl. *Consequence Index*) glavnog događaja izračunata je korištenjem FSA težinskog sustava:

$$C_i = \sum(share_i \times weight_i) \tag{18}$$

gdje su težine (engl. $weight_i$) definirane prema IMO FSA:

- 0,01 – male (engl. *Minor*),
- 0,1 – značajne (engl. *Significant),*
- 1 – teške (engl. *Severe*),
- 10 – katastrofalne (engl. *Catastrophic*).

Dok je udio pojedine razine posljedice (engl. s*hare*) definiran kao relativna frekvencija pojavljivanja te razine unutar promatranog glavnog događaja, izračunava se kao omjer broja pojavljivanja određene razine posljedice i ukupnog broja svih zabilježenih posljedica za taj događaj.

$$share_{i,j} = \frac{n_{i,j}}{\sum_j n_{i,j}} \tag{19}$$

gdje je:

- i – glavni događaj,
- j – razina posljedice (engl. *occurrence / severity level*),
- $n_{i,j}$ - broj pojavljivanja razine j unutar događaja i,
- $\sum_j n_{i,j}$ – ukupan broj svih slučajeva za taj događaj.

a za svaki glavni događaj vrijedi:

$$\sum_j \text{share}_{i,j} = 1 \tag{20}$$

Svaki glavni događaj i karakterizira skup varijabli koje doprinose njegovom nastanku. Svaka varijabla ima pridruženu vjerojatnost nastanka nezgode P_i dobivenu iz prediktivnog modela. Nakon toga se za svaku varijablu unutar pojedinog glavnog

događaja računa agregirana vjerojatnost P kao ponderirani prosjek pojedinačnih vjerojatnosti, pri čemu su ponderi određeni doprinosom riziku (R_i):

$$P = \frac{\sum_{i=1}^{n}(P_i \cdot R_i)}{\sum_{i=1}^{n} R_i} \tag{21}$$

gdje je:

- P_i– vjerojatnost varijable dobivena modelom,
- R_i– doprinos varijable riziku ($R_i = P_i \cdot C_i$),
- n– broj varijabli uključenih u nastanak glavnog događaja.

Važno je naglasiti da se kontinuirani rizik definiran izrazima (16) i (17) koristi za kvantitativnu analizu i procjenu doprinosa pojedinih varijabli, dok se za potrebe klasifikacije i vizualne interpretacije rizika primjenjuje diskretizirani pristup temeljen na FSA matrici rizika, u kojem se agregirana vjerojatnost (21) transformira u razine učestalosti (22), a konačna razina rizika određuje se kombinacijom diskretnih razina učestalosti i posljedica, čime se dobivaju cjelobrojne vrijednosti rizika unutar matrice.

Na temelju tako dobivenih vrijednosti omogućena je detaljna analiza doprinosa pojedinih varijabli ukupnom riziku, dok se agregirana razina rizika glavnog događaja interpretira kroz distribuciju i dominantne vrijednosti varijabli.

Nakon čega se P diskretizira u razinu učestalosti F prema FSA kategorijama:

$$F = f(P) = \begin{cases} F1 & \text{ako } 0 \le P < 0{,}01 \\ F2 & \text{ako } 0{,}01 \le P < 0{,}05 \\ F3 & \text{ako } 0{,}05 \le P < 0{,}1 \\ F4 & \text{ako } 0{,}1 \le P < 0{,}2 \\ F5 & \text{ako } P \ge 0{,}2 \end{cases} \tag{22}$$

Ovaj korak standardizira vjerojatnosti u diskretne razine učestalosti, što omogućava usporedbu i klasifikaciju rizika između različitih glavnih događaja. Empirijska frekvencija:

$$F_{emp} = \frac{n}{N} \tag{23}$$

gdje je:

- n= broj događaja,
- N= ukupno svih događaja.

Na temelju klasifikacije iz Tablice 4., gdje su definirane razine učestalosti (F1–F5) i posljedica (C1 – C4), ukupni rizik u matrici poprima vrijednosti u rasponu od 1 do 20.

Za razliku od empirijske frekvencije, u ovom radu frekvencija događaja određena je i kao agregirana vjerojatnost dobivena primjenom modela strojnog učenja. Vrijednost frekvencije izračunata je kao ponderirani prosjek pojedinačnih doprinosa varijabli, pri čemu su ponderi određeni njihovim doprinosom ukupnom riziku, čime se omogućuje integracija kompleksnih međusobnih odnosa između varijabli. Na temelju toga izračunao se parcijalni rizik po kategorijama varijabli (R_i), kao i ukupni rizik po varijablama.

Rezultati su kategorizirani na razine rizika (*Low, Medium, High*) primjenom kvantilne podjele podataka, radi omogućavanja relativne usporedbe među modelima.

Završne faze (15 – 18) obuhvaćale su grafičku i numeričku interpretaciju rezultata, njihovu pohranu i dokumentiranje te implementaciju ansambl i slojevitog pristupa radi povećanja stabilnosti i robusnosti modela. Cjelokupni analitički postupak dokumentiran je i pohranjen zajedno s pripadajućim R skriptama (RStudio Team, 2024; Kassambara, 2018), čime je osigurana potpuna znanstvena reproducibilnost i metodološka transparentnost istraživanja.

Svi rezultati analize, uključujući metričke vrijednosti, matrice zabune, predikcije modela, procjena vjerojatnosti nezavisnih varijabli izvezeni su u Excel format pomoću paketa *openxlsx*, čime je osigurana transparentnost i reproducibilnost provedenih postupaka. Također, kompletan programski kod napisan u R programskom okruženju, zajedno sa skriptama za obradu podataka, treniranje modela i vizualizaciju, sistematiziran je i priložen u posebnom prilogu na kraju doktorskog rada. Na taj način omogućena je naknadna provjera i potpuna replikacija analize u različitim istraživačkim kontekstima.
Na kraju svakog potpoglavlja koje obrađuje pojedinačnu komponentu metodološkog okvira, obradu podataka, selekciju značajki, treniranje i validaciju modela te evaluaciju rezultata, dodan je sažet pregled korištenih R tehnika i alata i primijenjenih funkcija. Ovaj tehnički pregled pruža jasan uvid u implementaciju svakog analitičkog koraka i dodatno doprinosi transparentnosti i razumljivosti provedenog istraživanja.

3.1. Izvor, kodiranje i opis ulaznih podataka

Primjena metoda umjetne inteligencije u procjeni i klasifikaciji rizika pomorskih nezgoda zahtijevala je temeljito pripremljene, strukturirane i kvalitetne ulazne podatke. Analiza se temeljila na informacijama o kretanju i položaju brodova, području plovidbe, tehničkim karakteristikama brodova, vremenskim uvjetima, gustoći prometa te podacima o ljudskim čimbenicima, uključujući iskustvo posade i povijest nezgoda. Pravilna priprema podataka uključivala je razvrstavanje odabranih numeričkih varijabli u unaprijed definirane intervale, kodiranje kategorija, pretvaranje ulaznih varijabli u faktorski oblik, standardizaciju numeričkih varijabli te nadomještanje nedostajućih vrijednosti, čime je skup podataka pripremljen za treniranje i evaluaciju modela. Primjena navedenih podataka omogućila je modelima potpornih vektora preciznu identifikaciju granica između klasa rizika, modelima slučajnih šuma procjenu važnosti varijabli i stabilnosti predikcija, dok su neuronske mreže uspješno prepoznale složene nelinearne odnose unutar podataka.

Provedena je detaljna analiza podataka o pomorskim nezgodama u obalnoj plovidbi, prikupljenih u razdoblju od 1. siječnja 2020. do 31. prosinca 2024. godine. Uključivanje nezgoda u analizu temeljilo se na završetku službene istrage, što je osiguravalo vjerodostojnost i potpunost podataka relevantnih za procjenu rizika. Podaci su korišteni isključivo u svrhu modeliranja i evaluacije rizika, u skladu s ciljevima istraživanja.

Najprije su kontaktirane nacionalne agencije koje se bave istraživanjem pomorskih nezgoda. Ured za istragu pomorskih nezgoda pri Ujedinjenom Kraljevstvu je dostavio tražene podatke putem zahtjeva za pristup informacijama (engl. *Freedom of Information*), pod referentnim brojevima FOI–00034036 i FOI–00031781. Osim Ujedinjenog Kraljevstva, ovdje spada i skupina država i teritorija koje koriste zastavu Ujedinjenog Kraljevstva (engl. *Red Ensign Group* – REG) i čiji pomorski registri djeluju u skladu s njegovim zakonodavstvom i međunarodnim standardima. Drugi subjekti, iako kontaktirani, nisu dostavili podatke, pa su ostale informacije prikupljene izravnim pretraživanjem službenih internetskih stranica relevantnih institucija. Konkretno, podaci su prikupljani i sistematizirani iz sljedećih izvora:

- GISIS (Global Integrated Shipping Information System, 2025),
- EMSA, putem platforme Europski informacijski sustav o pomorskim nezgodama (engl. *European Marine Casualty Information Platform* – EMCIP) (European Maritime Safety Agency, 2024),
- Australski ured za sigurnost prometa (engl. *Australian Transport Safety Bureau* – ATSB) (Australian Transport Safety Bureau, 2025),
- Američki nacionalni odbor za sigurnost prometa (engl. *National Transportation Safety Board* – NTSB) (National Transportation Safety Board, 2025),
- Komisija za istragu prometnih nezgoda Novog Zelanda (engl. *Transport Accident Investigation Commission* – TAIC) (Transport Accident Investigation Commission, 2025),
- Finska agencija za istraživanje nezgoda (engl. *Safety Investigation Authority of Finland* – SIAF) (Safety Investigation Authority of Finland, 2025),
- Danski odbor za pomorsku istragu nezgoda (engl. *Danish Maritime Accident Investigation Board* – DMAIB) (Danish Maritime Accident Investigation Board, 2025),
- Turski centar za istraživanje sigurnosti u prometu (engl. *Transport Safety Investigation Center*), (tur. *Ulaştırma ve Altyapı Bakanlığı Emniyet İnceleme Merkezi* – UEIM) (UEIM, 2025),
- Kanadski odbor za sigurnost u prometu (engl. *Transportation Safety Board of Canada* – TSB) (Transportation Safety Board of Canada. (TSBC), 2025),
- Odbor za pomorske nezgode Cipra (engl. *Marine Accident and Incident Investigation Committee* – MAIC) (Marine Accident and Incident Investigation Committee of the Republic of Cyprus, 2025),
- Savezni ured za istraživanje pomorskih nezgoda Njemačka (njem. *Bundesstelle für Seeunfalluntersuchung* – BSU) (Bundesstelle für Seeunfalluntersuchung, 2025),
- Državna komisija za istraživanje nezgoda Švedske (engl. *Swedish Accident Investigation Authority* – SAIA), (šved. *Statens haverikommission* – SHK) (Swedish Accident Investigation Authority, 2025).

U Jadranskom moru evidentirano je ukupno 11 pomorskih nezgoda (vidi Tablicu 12.), koje su bile predmet analize nadležnih agencija za istraživanje nezgoda u cilju unaprjeđenja sigurnosti plovidbe.

Prikupljeni podaci nisu bili unificirani, što je značilo da je pristup informacijama i njihova obrada varirala ovisno o državi i instituciji. U mnogim slučajevima bilo je nužno ručno analizirati pojedinačne izvještaje o nezgodama i iz njih ručno izdvajati i unositi informacije u Excel tablice. Kombinacija više baza podataka omogućila je sveobuhvatnu analizu, pri čemu je jedna baza dominantna zbog svoje detaljnosti i potpunosti, jer obuhvaća sve razine nezgoda, uključujući i događaje manjeg značaja, ostale baze obuhvaćaju isključivo veće nezgode, za koje su provedeni postupci analize i službene istrage. Kako bi se osigurala standardizacija i mogućnost strojnog procesuiranja podataka, varijable su definirane prema smjernicama iz Vodiča za javne korisnike Europskog informacijskog sustava o pomorskim nezgodama (engl. *European Marine Casualty Information Platform Guide for Public Users*) (European Maritime Safety Agency, 2024).

Pri tome su isključeni svi događaji koji se odnose na:

- usidrene brodove,
- požare,
- ozljede koje nisu posljedica događaja povezanih s glavnim pomorskim događajem,
- operacije traganja i spašavanja (engl. *Search and Rescue* – SAR).

U analizu su uključeni isključivo brodovi koji su bili u plovidbi.

Značajan doprinos u stručnoj obradi podataka dala su prethodna istraživanja o modeliranju nezgoda i procjeni rizika, koja su poslužila kao teorijska osnova za razvoj metodologije obrade i klasifikacije podataka (Antão et al., 2024; Bi et al., 2022; Cao et al., 2025; Zhang et al., 2025a).

Osim toga, korišteni su dostupni alati i programski jezici za analizu podataka, prvenstveno programsko okruženje za statistiku i obradu podataka (engl. *Integrated Development Environment* – RStudio) (RStudio Team, 2024), koje je omogućilo napredne statističke analize i pripremu podataka za izgradnju modela rizika.

Priprema podataka bila je sustavna i višefazna, obuhvaćajući ručno sređivanje, čišćenje, kodiranje i usklađivanje informacija iz različitih nacionalnih i međunarodnih izvora prema smjernicama Europske agencije za pomorsku sigurnost.

Ukupni skup podataka korišten u analizi sadrži 1591 slog, pri čemu svaki slog predstavlja jednu evidenciju rizične situacije u pomorskom prometu. *Glavni događaj* nije nužno *najvažnija* među svim varijablama, već je logičan izbor za ciljnu varijablu jer predstavlja ključni klasifikacijski ishod, dok ostale varijable služe kao ulazne značajke za predikciju. Ova varijabla klasificira sigurnosne događaje u sedam kategorija, označenih vrijednostima od 1 do 7, pri čemu svaka vrijednost odgovara određenoj vrsti *glavnog događaja* ili pomorske nezgode.

U metodološkom dijelu koji se odnosi na pripremu podataka, numeričke vrijednosti u takvim varijablama zamijenjene su odgovarajućim deskriptivnim oznakama radi poboljšanja interpretabilnosti i osiguranja pravilnog tretmana kategorija u kasnijim analitičkim postupcima. Transformacija se temeljila na unaprijed definiranom priručniku za kodiranje vrijednosti, tj. listi u kojoj je svakoj numeričkoj vrijednosti pridružena odgovarajuća tekstualna oznaka. Primjerice, u varijabli *baza podataka* vrijednost *1* zamijenjena je oznakom *MAIB*, a vrijednost *2* oznakom *GISIS*. Na isti način standardizirane su varijable povezane s bazama podataka, tipovima brodova, materijalima konstrukcije, vremenskim uvjetima, lokacijama događaja, fazama putovanja, prisutnošću peljara, zastavama brodova i drugim relevantnim čimbenicima.
Tablica 5. prikazuje pregled varijable ulaznog skupa podataka korištenih u analizi i klasifikaciji rizika pomorskih nezgoda. U tablici su navedeni redni broj, naziv svake varijable te kratko objašnjenje njezina značenja.

Većina varijabli definiranih u skupu podataka su faktorske (engl. *factor*) tj. kategorijske varijable, što znači da su imale unaprijed definirane klase koje se ponavljaju unutar uzorka. Međutim, tri varijable ulaznog skupa podataka *država izvjestiteljica*, *područje nezgode* i *zastava broda* definirane su kao znakovne (*nominalne*) (engl. *character*) varijable, jer se njihove vrijednosti sastoje od tekstualnih nizova koji mogu biti vrlo raznoliki i bez strogo ograničenog broja ponavljajućih kategorija.

Tablica 5. Opis svih varijabli ulaznog skupa podataka

Rb.	Varijabla	Prijevod (hrv.)	Značenje varijable
1	*main_event*	glavni događaj	Glavni događaj koji je doveo do pomorske nezgode
2	*database*	baza podataka	Izvor baze podataka iz koje potječe zapis
3	*occurrence*	ozbiljnost nezgode	Ozbiljnost pomorske nezgode
4	*year_reported*	godina prijave	Godina u kojoj je pomorska nezgoda prijavljena
5	*month*	mjesec	Mjesec u kojem je pomorska nezgoda prijavljena
6	*state_reporting*	država izvjestiteljica	Država koja je prijavila pomorsku nezgodu
7	*natural_light*	prirodno svjetlo	Uvjeti prirodnog osvjetljenja u trenutku pomorske nezgode (npr. dan, noć)
8	*sea_state*	stanje mora	Stanje mora u trenutku pomorske nezgode
9	*visibility*	vidljivost	Vidljivost u trenutku pomorske nezgode
10	*wind_force*	jačina vjetra	Jačina vjetra u trenutku pomorske nezgode
11	*weather*	vrijeme	Vremenski uvjeti (npr. magla, kiša, vedro)
12	*occurrence_location*	mjesto pomorske nezgode	Lokacija gdje se dogodila pomorska nezgoda (npr. prilazak luci, lučko područje)
13	*density*	gustoća prometa	Gustoća prometa
14	*location_area*	područje nezgode	Šire geografsko područje događaja (država)
15	*vessel_flag*	zastava broda	Zastava (država upisa broda)
16	*vessel_category*	kategorija broda	Kategorija broda prema vrsti upotrebe (npr. teretni, putnički)
17	*gt*	bruto tonaža	Bruto tonaža broda (engl. *gross tonnage*)
18	*vessel_type*	tip broda	Tip broda (npr. kontejnerski, brod za prijevoz tekućih tereta)
19	*length*	duljina	Duljina broda u metrima

20	*width*	širina	Širina broda u metrima
21	*damage*	oštećenje	Označava veličinu oštećenja broda
22	*pollution*	onečišćenje	Označava je li došlo do onečišćenja mora
23	*material*	materijal trupa	Označava materijal od kojeg je izgrađen brod
24	*ship_age*	starost broda	Starost broda u godinama
25	*vessel_sink*	potonuće broda	Označava je li brod potonuo
26	*crew*	posada	Broj članova posade na brodu
27	*voyage_segment*	segment putovanja	Dio putovanja tijekom kojeg se dogodila pomorska nezgoda (npr. ukrcaj, plovidba)
28	*pilot*	peljar	Označava je li na brodu bio prisutan lučki peljar
29	*injuries_total*	ukupno ozljeda	Ukupan broj ozlijeđenih osoba u pomorskoj nezgodi

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Varijable korištene u analizi pomorskih nezgoda podijeljene su u četiri temeljne skupine: *operativni čimbenici, okolišni čimbenici, karakteristike broda i posljedice*, dok je *vremensko-analitički* kontekst zasebno definiran. Takva podjela izvršena je radi strukturiranog pristupa klasifikaciji nezgoda, analizi i procjeni rizika, jer svaka skupina obuhvaća varijable koje odražavaju različite aspekte nezgode: uzroke, uvjete u kojima se događaj odvija, tehničke karakteristike broda i posljedice. Na temelju ove podjele moguće je: klasificirati nezgode u sedam skupina, analizirati važnost pojedinih čimbenika u nastanku nezgoda te procijeniti vjerojatnosti pojave glavnog događaja, posljedica događaja, ishoda unutar svake skupine, te konačno rizika.

U operativne čimbenike svrstane su varijable *segment putovanja*, *peljar*, *ozbiljnost nezgode* i *država izvjestiteljica*. One odražavaju fazu plovidbe, odnosno operativni kontekst u kojem se nezgoda dogodila. Takva grupacija omogućuje analizu kako različiti dijelovi putovanja, prisustvo peljara ili karakter samog događaja utječu na nastanak nezgoda te pridonosi pravilnoj klasifikaciji u odgovarajuće skupine rizika.

Okolišni čimbenici obuhvaćaju varijable koje odražavaju prirodne i prometne uvjete, uključujući *prirodno svjetlo*, *stanje mora*, *vidljivost*, *jačinu vjetra*, *vrijeme*, *mjesto pomorske nezgode*, *područje nezgode* i *gustoću prometa*. Ovi čimbenici imaju značajan utjecaj na vjerojatnost nastanka nezgode i njezinu ozbiljnost. Također, okolišni čimbenici su ključni za razvoj prediktivnih modela i izradu matrica zabune u klasifikaciji, jer omogućuju identificiranje uvjeta koji najčešće dovode do pogrešaka u klasifikaciji nezgoda ili procjeni većeg rizika.

Varijable vezane uz karakteristike broda, *bruto tonaža*, *zastava broda*, *kategorija broda*, *tip broda*, *duljina*, *širina*, *materijal trupa*, *starost broda* i *posada*, grupirane su radi analize tehničke i konstrukcijske ranjivosti broda. U ovu skupinu uključena je i varijabla *posada*, koja opisuje broj članova posade kao jedno od obilježja raspoloživih ljudskih resursa za obavljanje brodskih zadaća i reagiranje u izvanrednim situacijama. Njezino značenje potrebno je promatrati u odnosu na karakteristike broda i zahtjeve plovidbe, jer sama brojnost ne određuje osposobljenost ni učinkovitost posade. Iskustvo, znanje, uvježbanost i ljudske pogreške nisu izravno obuhvaćeni ovom varijablom, što je

u skladu s ograničenjima istraživanja navedenima u poglavlju 5.4. Ove varijable omogućuju razlikovanje rizika između različitih brodova i povećavaju točnost modela klasifikacijske i vjerojatnosti, jer određene karakteristike broda uvjetuju način i učestalost nastanka nezgoda te procjene većeg rizika.

U skupini posljedica nalaze se varijable *oštećenje*, *onečišćenje*, *ukupno ozljeda* i *potonuće broda*, koje kvantificiraju ishode pomorske nezgode te omogućuju definiranje razine ozbiljnosti događaja. Ove varijable koriste se za definiranje izlazne klase modela, dok model strojnog učenja procjenjuje vjerojatnost pripadnosti pojedinoj klasi. Na temelju tako dobivenih vjerojatnosti i pripadajućih razina ozbiljnosti može se interpretirati razina rizika, koja predstavlja analitičku kombinaciju vjerojatnosti i posljedica.

Varijable iz vremensko-analitičkog konteksta, *glavni događaj*, *baza podataka*, *godina prijave* i *mjesec*, koriste se za pozicioniranje događaja u vremenu i prema izvoru podataka. Iako nisu izravno uključene u klasifikaciju nezgoda, omogućuju analizu vremenskih obrazaca, praćenje trendova i usporedbu podataka između različitih baza.

Tablica 6. prikazuje strukturne karakteristike ulaznog skupa podataka o pomorskim nezgodama prije bilo kakve obrade ili transformacije. Njezin cilj je pružiti detaljan pregled svih ulaznih varijabli korištenih u klasifikaciji rizika, uključujući tip varijable (faktorska ili znakovna), broj razina, najčešću vrijednost (mod), učestalost pojavljivanja i broj nedostajućih vrijednosti. Takav pregled omogućava razumijevanje strukture ulaznog skupa, identifikaciju potencijalnih problema te planiranje daljnje pripreme podataka. Varijable su prvotno obrađene prema njihovom tipu i prirodi kako bi se olakšala obrada odnosno rukovanje podacima i omogućila sustavna priprema skupa za daljnju transformaciju i analizu, pri čemu su tekstualne varijable privremeno kodirane numeričkim oznakama radi sistematizacije i jednostavnije statističke provjere.
U tablici je prikazano ukupno 29 varijabli, a za svaku su navedeni sljedeći atributi: redni broj (Rb.), naziv varijable, tip podatka (faktorski ili tekstualni), broj razina (za faktorske varijable), najčešća vrijednost (mod), učestalost pojavljivanja najčešće vrijednosti u skupu (frenkevencija) te broj nedostajućih vrijednosti, kao i sve moguće razine koje faktorska/tekstualna varijabla može poprimiti.

Tablica 6. Strukturne karakteristike ulaznog skupa podataka

Rb.	Varijabla	Tip	Razina	Mod	Frekvencije	Bez vrijednosti	Moguće razine
1	*main_event*	faktor	7	5	370	0	1, 2, 3, 4, 5, 6, 7
2	*database*	faktor	13	1	1118	0	1, 2, 3, 5, 6, 7, 8 9, 10, 11, 12, 13
3	*occurrence*	faktor	4	4	970	0	1, 2, 3, 4
4	*year_reported*	faktor	5	4	522	0	1, 2, 3, 4, 5
5	*month*	faktor	12	10	160	0	1, 10, 11, 12, 2, 3, 4, 5, 6, 7, 8, 9
6	*state_reporting*	znak	25	UK	1118	0	25
7	*natural_light*	faktor	3	1	799	333	1, 2, 3

8	*sea_state*	faktor	9	4	224	657	1, 2, 3, 4, 5, 6, 7, 8, 9
9	*visibility*	faktor	5	4	573	724	1, 2, 3, 4, 5
10	*wind_force*	faktor	12	5	174	657	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13
11	*weather*	faktor	4	1	542	834	1, 2, 3, 4
12	*occurrence_location*	faktor	6	4	1085	0	1, 2, 3, 4, 5, 6
13	*density*	faktor	4	4	1440	0	1, 2, 3, 4
14	*location_area*	znak	37	UK	1080	0	37
15	*vessel_flag*	znak	36	UK	600	0	36
16	*vessel_category*	faktor	5	2	973	0	1, 2, 3, 4, 5
17	*gt*	faktor	5	1	597	0	1, 2, 3, 4, 5
18	*vessel_type*	faktor	13	2	490	0	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
19	*length*	faktor	4	1	1115	0	1, 2, 3, 4
20	*width*	faktor	3	1	1256	0	1, 2, 3
21	*damage*	faktor	4	4	764	272	1, 2, 3, 4
22	*pollution*	faktor	3	2	1175	293	1, 2, 3
23	*material*	faktor	5	4	1141	35	1, 3, 4, 5, 6
24	*ship_age*	faktor	5	5	497	336	1, 2, 3, 4, 5
25	*vessel_sink*	faktor	2	2	1204	330	1, 2
26	*crew*	faktor	5	1	594	707	1, 2, 3, 4, 5
27	*voyage_segment*	faktor	7	1	440	281	1, 2, 3, 4, 5, 6, 7
28	*pilot*	faktor	3	3	541	250	1, 2, 3
29	*injuries_total*	faktor	4	1	1388	153	1, 2, 3, 4

Izvor: izradio autor

Tablica 7. prikazuje raspodjelu kategorija ciljne varijable *glavni događaj*, koja klasificira sigurnosne događaje u sedam tipova nezgoda. Svaka vrijednost predstavlja specifičnu vrstu *glavnog događaja*, a broj pojavljivanja pokazuje koliko puta je ta nezgoda evidentirana u skupu podataka. Klasifikacija se temelji na standardiziranim pravilima Europskog informacijskog sustava o pomorskim nezgodama i prethodnim istraživanjima (Antão et al., 2024; Bi et al., 2022; Cao et al., 2025). U tablici su značenja svake numeričke vrijednosti ciljne varijable, prikazuje naziv klase na engleskom jeziku i odgovarajući prijevod na hrvatski. Time se osigurava jasno razumijevanje prirode svakog *glavnog događaja*, što je ključno za interpretaciju modelskih predikcija i povezivanje s realnim događajima u pomorskom prometu. Na primjer, vrijednost 2 označava *Sudar*, dok vrijednost 4 označava *Nasukavanje*. Prikazane su moguće vrijednosti ciljne varijable zajedno s njihovom učestalošću, odnosno brojem pojavljivanja u skupu podataka, prije zamjene nedostajućih vrijednosti (engl. *Not Available* ili *Missing Value* – NA). Uz navedenu učestalost, tablica olakšava razumijevanje prirode svake nezgode i omogućava

pravilnu interpretaciju podataka za modeliranje rizika. Klasifikacija omogućuje precizno praćenje izvora podataka i eventualno povezivanje karakteristika nezgoda s praksama izvještavanja pojedinih agencija ili regija.

Tablica 7. Značenja vrijednosti varijable *glavni događaj* (*main_event*)

Vrijednost	Kategorija (engleski)	Prijevod na hrvatski	Broj pojavljivanja
1	*Capsizing / Listing*	Prevrtanje / Naginjanje	80
2	*Collision*	Sudar	350
3	*Contact*	Udaranje u nepomični objekt (bez sudara s drugim plovilom)	277
4	*Grounding / Stranding*	Nasukavanje	322
5	*Damage / Loss of equipment*	Oštećenje / Gubitak opreme	370
6	*Flooding / Foundering*	Prodor mora / Potonuće	90
7	*Loss of control*	Gubitak upravljivosti	102

Izvor: izradio autor prema (EMSA, 2024)

Varijabla *baza podataka* Tablica 8. označava različite izvore iz kojih su prikupljeni podaci o događajima, pomorskim nezgodama ili zapisima koji su analizirani u okviru istraživanja. Svaka vrijednost u stupcu *baza podataka* predstavlja kraticu naziva nacionalne ili međunarodne agencije ili tijela za istraživanje pomorskih nezgoda. Varijabla identificira izvor podataka ili bazu iz koje je pojedinačna evidencija o nezgodi preuzeta. Broj pojavljivanja jasno pokazuje učestalost zapisa po svakom izvoru, što omogućava uvid u dominirajuće baze podataka u skupu i pruža kontekst za analizu raspodjele podataka prije obrade i modeliranja.

Tablica 8. Značenje i broj pojavljivanja vrijednosti varijable *baza podataka* (*database*)

Vrijednost	Baza podataka	Puni naziv baze / agencije	Prijevod na hrvatski	Broj pojavljivanja
1	MAIB	*Marine Accident Investigation Branch (UK)*	Ured za istraživanje pomorskih nezgoda (Ujedinjeno Kraljevstvo – UK)	1118
2	GISIS	*Global Integrated Shipping Information System*	Globalni integrirani informacijski sustav za pomorstvo (IMO)	89
3	NTSB	*National Transportation Safety Board (USA)*	Nacionalni odbor za sigurnost prometa (SAD)	147
4	REG	*Red Ensign Group*	Skupina zastava pod jurisdikcijom UK–a	2
5	TAIC	*Transport Accident Investigation Commission (NZ)*	Komisija za istragu prometnih nezgoda (Novi Zeland)	6
6	TSB	*Transportation Safety Board (Canada)*	Odbor za sigurnost prometa (Kanada)	189
7	ATSB	*Australian Transport Safety Bureau*	Australski ured za sigurnost prometa	9
8	SIAF	*Safety Investigation Authority of Finland*	Agencija za istraživanje nezgoda (Finska)	5

9	BSU	*Federal Bureau of Maritime Casualty Investigation (Germany)*	Savezni ured za istraživanje pomorskih nezgoda (Njemačka)	8
10	SAIA	*Swedish Accident Investigation Authority*	Švedska agencija za istraživanje nezgoda	5
11	DMAIB	*Danish Maritime Accident Investigation Board*	Danski odbor za istraživanje pomorskih nezgoda	6
12	MAIC	*Marine Accident and Incident Investigation Committee (Cyprus)*	Odbor za pomorske nezgode i incidente (Cipar)	2
13	UEIM	*Ulaştırma ve Altyapı Bakanlığı Emniyet İnceleme Merkezi (Turkey)*	Turski centar za istraživanje sigurnosti u prometu	5

Izvor: izradio autor

Varijabla *ozbiljnost nezgode* kategorizira ozbiljnost nezgode prema prethodno definiranim vrijednostima od 1 do 4 (Tablica 9.). U provedenoj retrospektivnoj klasifikaciji korištena je kao ulazna varijabla modela, a njezina raspodjela po vrstama glavnog događaja poslije je korištena za izračun indeksa posljedica C_i. Budući da je ozbiljnost poznata tek nakon događaja, ta varijabla nije ulaz za predviđanje prije njegova nastanka. Vrijednost 1 označava vrlo ozbiljne nezgode, dok vrijednost 4 označava nezgode klasificirane kao pomorski incident, koji su po intenzitetu najmanje ozbiljni. Tablica prikazuje ove klase s opisima na engleskom jeziku i njihovim prijevodom na hrvatski, čime se omogućava jasna interpretacija ozbiljnosti nezgoda u skupu podataka. U tablici su prikazane sve vrijednosti varijable *ozbiljnost nezgode* s brojem njihovih pojavljivanja u skupu podataka. Iz tablice je vidljivo da je najčešća vrijednost 4 (970 pojavljivanja), dok su vrijednosti 1 i 2 zastupljene jednako (po 163), a vrijednost 3 se pojavljuje 295 puta. Ova distribucija pokazuje da su određene klase događaja znatno češće od drugih.

Tablica 9. Klasifikacija i broj pojavljivanja varijable *ozbiljnost nezgode* (*occurrence*)

occurrence	**Opis (engleski)**	**Prijevod na hrvatski**	**Broj pojavljivanja**
1	*Very serious*	Vrlo teške posljedice	163
2	*Serious*	Nezgode s ozbiljnim posljedicama	163
3	*Less serious*	Nezgode manjeg intenziteta	295
4	*Marine incident*	Pomorski incident	970

Izvor: izradio autor prema (EMSA, 2024)

Tablica 10. prikazuje vrijednosti varijable *godina izvještavanja* u kojoj je određeni događaj ili zapis prijavljen. Tablica prikazuje broj pojavljivanja svake od tih vrijednosti u skupu podataka. Najviše nezgoda evidentirano je 2023. godine (vrijednost 4, 522 slučaja), dok je najmanje zabilježeno u 2021. godini (vrijednost 2, 83 slučaja). Ova tablica omogućavaju uvid u raspodjelu događaja kroz promatrano razdoblje i služe isključivo kao opisni prikaz podataka.

Relativno manji broj prijava u 2020. i 2021. godini mogao bi biti djelomično posljedica globalne pandemije bolesti Covid-19, koja je poremetila brojne administrativne i operativne procese, uključujući prikupljanje i prijavu podataka.

Tablica 10. Vrijednosti i broj pojavljivanja varijable *godina izvještavanja* (*year_reported*)

Vrijednost	*godina izvještavanja (year_reported)*	Broj pojavljivanja
1	2020	84
2	2021	83
3	2022	439
4	2023	522
5	2024	463

Izvor: izradio autor

Tablica 11. prikazuje distribuciju pojavljivanja varijable *mjesec* u skupu podataka. Najveći broj nezgoda zabilježen je u listopadu (vrijednost 10, 160 slučajeva), dok je najmanji zabilježen u prosincu (vrijednost 12, 100 slučajeva), čime se opisuje vremenska raspodjela događaja unutar promatranog razdoblja. Budući da analizirani skup podataka obuhvaća zapise s obje hemisfere, 96,9 % odnosi se na sjevernu hemisferu, sezonska interpretacija (ljeto – jesen) ne uvodi metodološku pogrešku te je varijabla vremenske raspodjele po mjesecima zadržana kao relevantna za analizu.

Tablica 11. Broj prijavljenih pomorskih nezgoda prema *mjesecu* (*month*)

Vrijednost	Broj pojavljivanja
1	142
2	115
3	102
4	136
5	132
6	123
7	144
8	157
9	139
10	160
11	141
12	100

Izvor: izradio autor

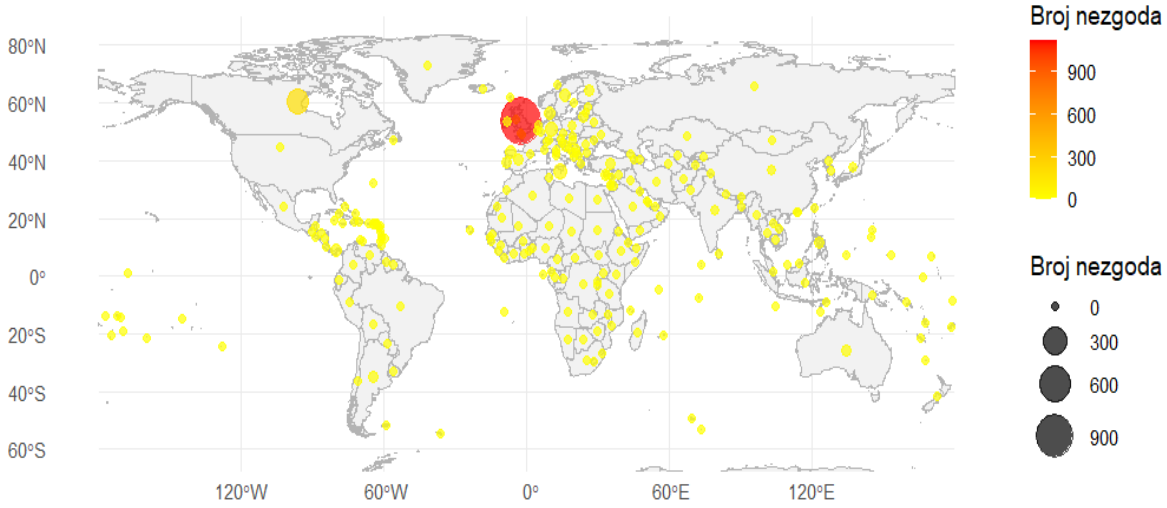
Tablica 12. prikazuje broj prijavljenih pomorskih nezgoda po *državi izvjestiteljici*. Najviše nezgoda evidentirano je za Ujedinjeno Kraljevstvo (1118), što odražava obim i detaljnost Ureda za istraživanje nezgoda pri Ujedinjenom Kraljevstvu, te ima ključnu ulogu u pouzdanosti i reprezentativnosti analize rizika u obalnoj plovidbi. Važno je napomenuti da su tijela Ujedinjenog Kraljevstva bila posebno transparentna te su dostavila cjelovita i točna izvješća, što dodatno objašnjava njihovu dominaciju u prikupljenim podacima.

Tablica 12. Broj prijavljenih pomorskih nezgoda prema *državi izvjestiteljici* (*state_reporting*)

Vrijednost	Broj pojavljivanja
ARGENTINA	11
AUSTRALIA	10
CANADA	198
CHINA	4
CROATIA	11
CYPRUS	6
DENMARK	17
FINLAND	7
FRANCE	13
GERMANY	24
INDIA	4
IRELAND	4
JAPAN	4
LITHUANIA	6
MALTA	41
NEW_ZEALAND	10
PANAMA	17
PHILIPPINES	5
POLAND	4
PORTUGAL	11
SPAIN	9
SWEDEN	12
TURKEY	5
UK	1118
UNITED_STATES	40

Izvor: izradio autor

Slika 6. prikazuje globalnu distribuciju prijavljenih pomorskih nezgoda po državama izvjestiteljicama. Veličina i boja točaka proporcionalni su broju nezgoda, pri čemu veće i crvenije točke označavaju veći broj nezgoda. Visoka zastupljenost Ujedinjenog Kraljevstva u analiziranoj bazi može odražavati obuhvat i detaljnost izvještavanja korištenog izvora te ne znači nužno da je u njegovim obalnim vodama najveći broj nezgoda ili najviša razina rizika, kako je detaljnije objašnjeno u nastavku.



Slika 6. Globalna distribucija pomorskih nezgoda

Izvor: izradio autor

Tablica 13. prikazuje klasifikaciju varijable *prirodno osvjetljenje* zajedno s njihovim opisima na engleskom i prijevodom na hrvatski jezik. Ova tablica omogućava jasno razumijevanje uvjeta osvjetljenja tijekom pomorskih nezgoda, što je važno za analizu utjecaja svjetlosnih uvjeta na sigurnost plovidbe. Tablica prikazuje broj pojavljivanja svake klase varijable *prirodno osvjetljenje* u skupu podataka. Iz tablice je vidljivo da je najčešće prijavljeno dnevno svjetlo (799 slučajeva), dok je 399 nezgoda zabilježeno noću, a 60 u sumraku ili zori. Nedostajuće vrijednosti (*NA*) iznose 333 slučaja, što ukazuje na nekompletne podatke za određene nezgode. Ova raspodjela pruža uvid u uvjete osvjetljenja tijekom pomorskih nezgoda, naglašava važnost obrade nedostajućih podataka i doprinosi pravilnoj pripremi skupa za modeliranje rizika.

Tablica 13. Klasifikacija i broj pojavljivanja varijable *prirodno osvjetljenje* (*natural_light)*

natural_light	**Opis (engleski)**	**Prijevod na hrvatski**	**Broj pojavljivanja**
1	*Daylight*	Dnevna svjetlost	799
2	*Twilight*	Sumrak/zora	60
3	*Night*	Noć	399
NA	*Unknown*	Nepoznato	333

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Prema podacima Državnog hidrometeorološkog zavoda (DHMZ), opis stanja za svaku razinu iz Douglasove ljestvice je prikazan Tablicom 14. (DHMZ, 2025).

Douglasova ljestvica (engl. *Douglas Sea Scale*) standardizirana je i prepoznata u međunarodnim meteorološkim praksama kao pouzdana mjera za komunikaciju stanja mora. Vrijednosti od 1 do 10 označavaju različite stupnjeve stanja mora, pri čemu niži brojevi znače mirnije more, a viši brojevi označavaju jače valove i teže uvjete na moru. Vrijednost *NA* označava poseban ili neklasificirani tip stanja mora (npr. *nepoznato*, što je

najčešće zabilježena vrijednost u ovom skupu podataka. Douglas-ova ljestvica stanja mora standardna je metoda ocjene stanja morske površine prema visini valova, izražena u deset razreda od 0 do 9. Ova ljestvica koristi se u pomorskim navigacijskim prognozama kako bi se opisala uvjetna valovitost i pogled na morsku površinu, a unutar ovog skupa podataka predstavljeni su aspekti te ljestvice (1–10). U tablici su također prikazane vrijednosti broj pojavljivanja, koja označava stanje mora u trenutku kada je došlo do pomorske nezgode. Svaka brojčana vrijednost odgovara određenom stupnju stanja mora prema Douglas-ovoj ljestvici, a broj pojavljivanja pokazuje koliko je puta to stanje zabilježeno u skupu podataka.

Tablica 14. Douglas-ova ljestvica i broj pojavljivanja varijable *stanja mora* (*sea_state*)

Vrijednost	Stanje mora	Visina valova (m)	Kratki opis izgleda mora	Broj pojavljivanja
1	0 – mirno (glatko, zrcalno, bonaca)	≈ 0	more poput zrcala	99
2	1 – mirno (naborano)	0 – 0.1	mali valići ili bore s pojavama	217
3	2 – malo valovito (valićasto)	0.1 – 0.5	kratki ili mali valovi; bregovi izgledaju staklasto	194
4	3 – umjereno valovito	0.5 – 1.2	veći valovi, bjeline, šuštanje	224
5	4 – valovito	1.25 – 2.5	valovi s mnogo bjelina, prskanje	108
6	5 – jače valovito	2.5 – 4	valovi se propinju, neprekidne bjeline	39
7	6 – uzburkano	4 – 6	veliki valovi s gustim prugama pjene	21
8	7 – teško	6 – 9	valovi se ruše, pjenušavi bregovi smanjuju vidljivost	5
9	8 – vrlo teško	9 – 14	iznimno veliki valovi, gusti morski dim	27
10	9 – izuzetno teško	> 14	složeni valovi, interferencija, ekstremno stanje	0
NA	nepoznato	–	stanje nepoznato ili nedovoljno podataka	657

Izvor: izradio autor prema (DHMZ, 2025)

Tablica 15. jasno prikazuje klasifikaciju i učestalost pojavljivanja različitih razina vidljivosti u pomorskim nezgodama. Tablica definira klasifikaciju vidljivosti od vrlo loše do vrlo dobre, što omogućava standardiziranu interpretaciju podataka. Prijevod na hrvatski je precizan i zadržava jasnoću engleskih kategorija. Analiza učestalosti pokazuje da se najveći broj nezgoda dogodio pri dobroj do vrlo dobroj vidljivosti, dok je vrlo loša vidljivost (vrijednost 1) zabilježena rijetko. To može upućivati na paradoksalni zaključak da većina nezgoda ne nastaje nužno u ekstremno nepovoljnim uvjetima vidljivosti, već i kada su uvjeti relativno povoljni, što sugerira ulogu drugih čimbenika u sigurnosti plovidbe. Ukupno, raspodjela podataka ukazuje na raznovrsne uvjete vidljivosti pri nezgodama, a kombinacija tablica omogućava razumijevanje kako kvantitativnih vrijednosti, tako i kvalitativnog opisa uvjeta vidljivosti.

Tablica 15. Klasifikacija i broj pojavljivanja varijable *vidljivost* (*visibility)*

visibility	**Opis (engleski)**	**Prijevod (hrvatski)**	**Broj pojavljivanja**
1	*Very poor – Vis < 0.5 NM*	*Vrlo loša – vidljivost < 0.5 Nautičke milje (engl. Nautical miles – NM)*	5
2	*Poor – 0.5 ≤ Vis < 2.0 NM*	*Loša – 0.5 ≤ vidlj. < 2.0 NM*	77
3	*Moderate – 2.0 ≤ Vis < 5.0 NM*	*Umjerena – 2.0 ≤ vidlj. < 5.0 NM*	86
4	*Good – 5.0 ≤ Vis < 25.0 NM*	*Dobra – 5.0 ≤ vidlj. < 25.0 NM*	573
5	*Very good – Vis ≥ 25.0 NM*	*Vrlo dobra – vidlj. ≥ 25.0 NM*	126
NA	*Unknown*	*Nepoznato*	724

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Tablica 16. prikazuje klasifikaciju varijable *jačina vjetra* prema Beaufortovoj ljestvici i raspodjelu pojavljivanja pojedinih razina vjetra u pomorskim nezgodama. Tablica definira standardizirane opise jačine vjetra od tišine do orkanskog vjetra, uključujući i vrijednost *NA* za nepoznate ili netipične podatke, a prijevod na hrvatski precizno odražava engleske klase i pokazuje da se najveći broj nezgoda dogodio pri umjerenom do slabom vjetru (vrijednosti 3–5), dok ekstremne vrijednosti (10 i 12) pojavljuju se rijetko. Vrijednosti 11 i 13 nisu zabilježene u skupu podataka, no njihovo zadržavanje u tablici važno je radi dosljednosti i cjelovitosti klasifikacijske sheme.
Ova raspodjela pokazuje da većina pomorskih nezgoda nastaje u uvjetima umjerenog vjetra, što ukazuje da i pri relativno povoljnim uvjetima postoji značajan rizik, dok ekstremni vjetrovi rjeđe uzrokuju zabilježene nezgode.

Tablica 16. Klasifikacija i broj pojavljivanja varijable *jačina vjetra* (*wind_force*) prema Beaufortovoj ljestvici

Vrijednost	**Opis (engleski)**	**Prijevod (hrvatski)**	**Broj pojavljivanja**
1	*0 – Calm – knot (0–1) m/s (0–1)*	0 – Tišina – čvor (0–1) / m/s (0–1)	35
2	*1 – Light air – knot (1–3) m/s (1–2)*	1 – Lahor – čvor (1–3) / m/s (1–2)	115
3	*2 – Light Breeze – knot (4–6) m/s (2–3)*	2 – Slab povjetarac – čvor (4–6) / m/s (2–3)	167
4	*3 – Gentle Breeze – knot (7–10) m/s (4–5)*	3 – Povjetarac – čvor (7–10) / m/s (4–5)	170
5	*4 – Moderate Breeze – knot (11–16) m/s (6–8)*	4 – Slab vjetar – čvor (11–16) / m/s (6–8)	174
6	*5 – Fresh Breeze – knot (17–21) m/s (9–11)*	5 – Umjeren vjetar – čvor (17–21) / m/s (9–11)	92
7	*6 – Strong Breeze – knot (22–27) m/s (11–14)*	6 – Umjereno jak vjetar – čvor (22–27) / m/s (11–14)	80

8	*7 – Near Gale – knot (28–33) m/s (14–17)*	7 – Jak vjetar – čvor (28–33) / m/s (14–17)	48
9	*8 – Gale – knot (34–40) m/s (17–21)*	8 – Umjerena oluja – čvor (34–40) / m/s (17–21)	16
10	*9 – Strong Gale – knot (41–47) m/s (21–24)*	9 – Umjereno jaka oluja – čvor (41–47) / m/s (21–24)	32
11	*10 – Storm – knot (48–55) m/s (25–28)*	10 – Jaka oluja – čvor (48–55) / m/s (25–28)	0
12	*11 – Violent Storm – knot (56–63) m/s (29–32)*	11 – Orkanska oluja – čvor (56–63) / m/s (29–32)	5
13	*12 – Hurricane – knot (+64) m/s (+33)*	12 – Orkan – čvor (64+) / m/s (33+)	0
NA	*UNKNOWN – 14*	Nepoznato / vrijednost izvan standardne ljestvice, moguće kodiranje greškom	657

Izvor: izradio autor prema (NOAA, 2025)

Tablica 17. prikazuju klasifikaciju varijable *vremenski uvjeti* i učestalost pojavljivanja različitih vremenskih uvjeta u pomorskim nezgodama. Tablica jasno definira klase od vedrog vremena do snijega, uključujući i nepoznate podatke (*NA*), a prijevod na hrvatski precizno odražava engleske opise. Analiza učestalosti pokazuje da se većina nezgoda dogodila pri vedrom vremenu (542 slučaja) i naoblačenom vremenu (135 slučajeva), dok su magla (13 slučajeva) i kiša (67 slučajeva) znatno rjeđi uvjeti, a snijeg nije zabilježen u obrađenom skupu podataka. Ova raspodjela sugerira da se većina pomorskih nezgoda dogodila u relativno povoljnim vremenskim uvjetima, što upućuje na važnost drugih čimbenika u nastanku nezgoda.

Tablica 17. Klasifikacija i broj pojavljivanja varijable *vremenski uvjeti* (*weather*)

Vrijednost	Opis (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Clear*	Vedro	542
2	*Fog*	Magla	13
3	*Overcast*	Naoblačeno	135
4	*Rain*	Kiša	67
5	*Snow*	Snijeg	0
NA	*Unknown*	Nepoznato	834

Izvor: izradio autor

Tablica 18. prikazuje klasifikaciju i broj pojavljivanja varijable *mjesto pomorske nezgode* s opisima na engleskom i njihovim prijevodom na hrvatski, kao i broj pojavljivanja. Od ukupno 1591 zapisa, 1085 se odnosi na lučko područje, 355 na prilaz luci, 77 na tjesnac, 54 na kanal, 11 na arhipelaški plovni put, a devet na ostale lokacije. Najviše nezgoda zabilježeno je u lučkim područjima (1085), dok su ostale klase, poput arhipelaških plovnih putova i tjesnaca, zabilježile znatno manji broj nezgoda.

Tablica 18. Klasifikacija i broj pojavljivanja varijable *mjesto pomorske nezgode* (*occurrence_location*)

Vrijednost	Opis (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Port approach*	Prilaz luci	355
2	*Archipelago fairway*	Arhipelaški plovni put	11

3	*Channel*	Kanal	54
4	*Port area*	Lučko područje	1085
5	*Strait*	Tjesnac	77
6	*Other*	Ostalo	9

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Tablica 19. prikazuje klasifikaciju i broj pojavljivanja varijable *gustoća prometa* s opisima na engleskom i hrvatskom jeziku. Gustoća prometa je podijeljena u četiri klase: niska, srednja, povišena i visoka gustoća, pri čemu se klasifikacija temelji na analizama podataka o plovidbenim rutama i prometu u prethodnim istraživanjima (Wu et al., 2016; Kim et al., 2022; Mou et al., 2024), ali je u ovom radu dodatno metodološki razrađena.

U literaturi se gustoća pomorskog prometa najčešće izražava kao broj plovila po satu i jedinici površine (brod/sat/km²), pri čemu se razlikuju osnovne razine: niska (do 1 brod/sat/km²), srednja (1–3 broda/sat/km²) i visoka (>3 broda/sat/km²) (Hasanspahić et al., 2019; PIANC, 1992). U skladu s tim, u ovom radu uvedena je dodatna kategorija povišene gustoće kao prijelazna razina između srednje i visoke gustoće, uvažavajući nelinearan porast rizika s povećanjem gustoće prometa. Povišena gustoća nije definirana kao fiksni raspon, već kao prijelazna zona u kojoj rizik počinje naglo rasti, što je u skladu s literaturom koja upućuje na nelinearan rast rizika, s ciljem preciznijeg modeliranja njegova postupnog povećanja (Feixiang & Lei, 2013).

S obzirom na ograničenu dostupnost izravnih kvantitativnih podataka o gustoći prometa za svaku pojedinu nezgodu, pripadnost nezgode određenoj kategoriji gustoće određena je na temelju lokacije događaja i tipičnih prometnih karakteristika područja. Tako su nezgode u lučkim područjima i prilazima lukama klasificirane kao visoka gustoća, nezgode u kanalima i tjesnacima kao povišena gustoća, nezgode u arhipelaškim i obalnim plovnim putovima kao srednja gustoća, dok su nezgode u otvorenijim i manje frekventnim područjima svrstane u kategoriju niske gustoće. Ovakav pristup omogućuje konzistentnu i metodološki utemeljenu dodjelu kategorije gustoće za svaku pojedinu nezgodu, uz istovremeno uvažavanje ograničenja dostupnih podataka.
Najviše nezgoda zabilježeno je u uvjetima visoke gustoće prometa (1440), dok su niska i srednja gustoća zabilježile znatno manji broj nezgoda.

Tablica 19. Klasifikacija i broj pojavljivanja varijable *gustoća prometa* (*density*)

Vrijednost	Opis (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Low density*	Niska gustoća	9
2	*Moderate density*	Srednja gustoća	11
3	*Elevated density*	Povišena gustoća	131
4	*High density*	Visoka gustoća	1440

Izvor: izradio autor prema (Wu et al., 2016; Kim et al., 2022; Mou et al., 2024)

Tablica 20. prikazuje raspodjelu vrijednosti varijable *područje nezgode* (37 mogućih), koja označava zemljopisno područje u kojem je zabilježena pomorska nezgoda. Varijabla je kvalitativnog (nominalnog) karaktera, a broj pojavljivanja prikazuje učestalost pomorskih nezgoda po državama ili regijama.

Najveći broj zabilježenih nezgoda odnosi se na Ujedinjeno Kraljevstvo, s ukupno 1080 slučajeva, što čini izrazito dominantan udio u skupu podataka zbog svoje detaljnosti i potpunosti, obuhvaćene su sve razine nezgoda, uključujući i događaje manjeg značaja. Važno je napomenuti da ova dominacija ne mora nužno odražavati stvarni broj nezgoda unutar obalnih voda UK, već je povezana s administrativnom odgovornošću ili zastavom broda, budući da britanska nadležna tijela istražuju i nezgode brodova koji plove pod zastavom UK-a, bez obzira na to gdje se nezgoda dogodila.

Značajniji broj slučajeva također je zabilježen u kanadskim provincijama, uključujući British Columbiu (62), Ontario (31), Nova Scotia (29), Quebec (41) i Newfoundland and Labrador (20), što ukazuje na relativno učestale nezgode u tim obalnim područjima. Među ostalim državama s većim brojem pomorskih nezgoda ističu se Malta (42), Njemačka (22), Francuska (19) i Danska (17), dok su ostale zemlje, poput Belgije, Hong Konga, Filipina i Prince Edward Islanda, zabilježene s vrlo malim brojem slučajeva (manje od 5).

Ova raspodjela ukazuje na značajne razlike u geografskoj učestalosti pomorskih nezgoda, koje mogu biti posljedica različitih čimbenika: intenziteta pomorskog prometa, geografskih i klimatskih uvjeta, organizacije sigurnosnog sustava te obveze i kvalitete izvještavanja u pojedinim zemljama. S obzirom na potencijalnu administrativnu dominaciju UK u izvještajima, interpretaciju ovih podataka treba promatrati s oprezom, kako se ne bi pogrešno zaključivalo da UK ima stvarno najveći broj nezgoda.

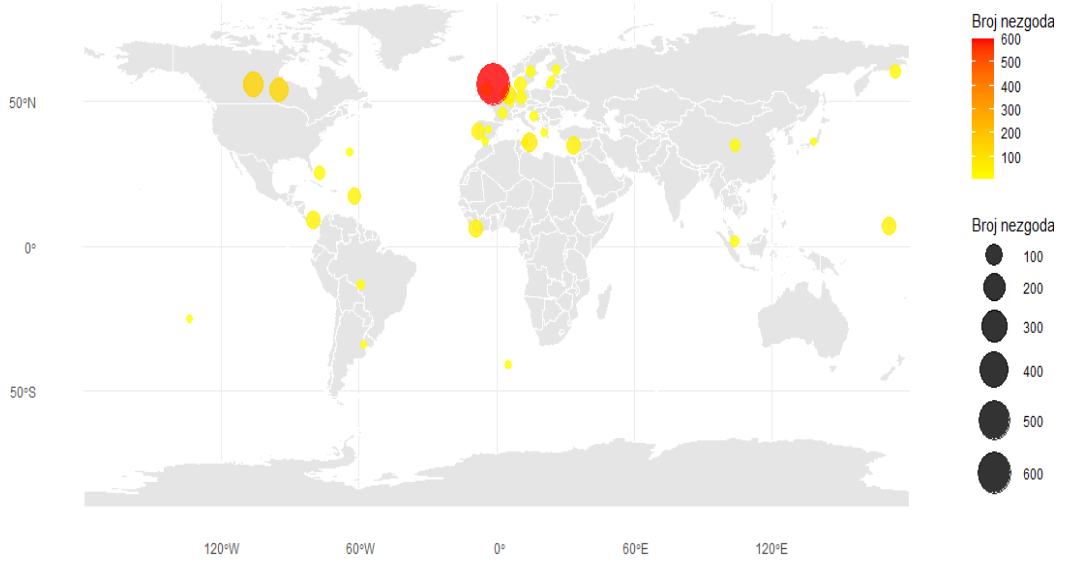
Tablica 20. Vrijednosti varijable *područje nezgode* (*location_area*)

Vrijednost	Broj pojavljivanja
ARGENTINA	16
AUSTRALIA	15
BELGIUM	1
BRITISH_COLUMBIA	62
CHINA	6
CROATIA	11
CYPRUS	9
DENMARK	17
FINLAND	5
FRANCE	19
GERMANY	22
HONG_KONG	4
IRELAND	11
ITALY	5
JAPAN	6
LIBERIA	11
LITHUANIA	6
MALTA	42
NETHERLANDS	6
NEW_BRUNSWICK	6

NEW_ZEALAND	10
NEWFOUNDLAND_AND_LABRADOR	20
NORWAY	7
NOVA_SCOTIA	29
ONTARIO	31
OUTSIDE_PROVINCIAL_BOUNDARIES	6
PANAMA	27
PHILIPPINES	3
POLAND	4
PORTUGAL	11
PRINCE_EDWARD_ISLAND	4
QUEBEC	41
SINGAPORE	6
SPAIN	13
SWEDEN	14
TURKEY	5
UK	1080

Izvor: izradio autor

Slika 7. grafički prikazuje raspodjelu pomorskih nezgoda prema *zastavi broda*, čime se vizualno naglašavaju države čije su zastave najčešće zastupljene u nezgodama. Najzastupljenije su zastave Ujedinjenog Kraljevstva, Sjedinjenih Američkih Država i Kanade, dok su zastave manjih flota ili država s manjim obalnim prometom prikazane s vrlo niskim frekvencijama. Slika jasno odražava strukturu podataka iz Tablice 21. i olakšava uočavanje razlika u učestalosti nezgoda prema državi zastave. Zbog preklapanja simbola, karta služi orijentacijskom prikazu zastupljenosti zastava brodova u analiziranoj bazi. Simboli ne označavaju lokacije nezgoda, a točan broj zapisa za svaku zastavu prikazan je u pripadajućoj tablici.



Slika 7. Pomorske nezgode prema zastavi broda

Izvor: izradio autor

Tablica 21. Vrijednosti varijable *zastava broda* (*vessel_flag*)

Vrijednost	Broj pojavljivanja
ANTIGUA_AND_BARBUDA	45
ARGENTINA	4
AUSTRALIA	4
BAHAMAS	24
BARBADOS	10
BELGIUM	8
BERMUDA	6
CANADA	148
CHINA	16
CROATIA	10
CYPRUS	55
DENMARK	32
FINLAND	10
FRANCE	11
GERMANY	17
GIBRALTAR	6
GREECE	5
ISLE_OF_MAN	20
ITALY	5
JAPAN	4
LATVIA	6
LIBERIA	54
LITHUANIA	4
LUXEMBOURG	4
MALTA	73
MARSHALL_ISLANDS	55
NETHERLANDS	62
NEW_ZEALAND	6
NORWAY	28
PANAMA	51
PORTUGAL	43
SINGAPORE	14
SPAIN	4
SWEDEN	11
UK	600
UNITED_STATES	136

Izvor: izradio autor

Tablica 22. prikazuje popis kategorija i broj pojavljivanja varijable *kategorija broda*. Vrijednost 1 odnosi se na ribarske brodove, vrijednost 2 na trgovačke brodove od najmanje 100 BT, vrijednost 3 obuhvaća razne druge klase, vrijednost 4 označava rekreacijska nekomercijalna plovila, dok vrijednost 5 predstavlja mala komercijalna plovila ispod 100 BT. Tablica daje jasan pregled klasifikacije korištene u analizi i broj

pojavljivanja svake od kategorija. Kategorija 2 (trgovački brodovi ≥100 BT) najzastupljenija je s 973 slučaja. Slijede kategorija 5 sa 282 pojavljivanja, kategorija 1 sa 174, te kategorija 4 sa 135 pojavljivanja. Najmanje je zastupljena kategorija 3 s ukupno 27 pojavljivanja. Tablica omogućuje uvid u distribuciju podataka po klasama.

Tablica 22. Klasifikacija i broj pojavljivanja varijable *kategorija broda* (*vessel_category*)

Vrijednost	Kategorija broda (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Fishing*	Ribarski brodovi	174
2	*Merchant ≥100 GT*	Trgovački brodovi ≥100 BT	973
3	*Other – assorted*	Ostalo – razne klase	27
4	*Recreational (non commercial)*	Rekreacijski brodovi (nekomercijalna)	135
5	*Small commercial (under 100 GT)*	Mali komercijalni brodovi (<100 BT)	282

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Tablica 23. daje klasifikaciju i broj pojavljivanja varijable *bruto tonaže broda* (engl. *gross_tonnage* – GT), pri čemu su vrijednosti raspoređene u pet raspona BT. Svaka kategorija opisuje tipične vrste brodova koje joj pripadaju: najmanji raspon (0 – 99 BT) obuhvaća uglavnom male rekreacijske i ribarske brodove; sljedeća skupina (100 – 199,99 BT) obuhvaća manje trgovačke brodove i servisna plovila; kategorija 3 obuhvaća brodove u rasponu 200 – 499,99 BT, među kojima su najčešće manji teretni ili multifunkcionalni komercijalni brodovi; kategorija 4 (500 – 2999,99 BT) uključuje srednje do veće trgovačke brodove, poput RO – RO putničkih brodova i manjih kontejnerskih brodova; dok kategorija 5 označava brodove od 3000 BT naviše, tipične za međunarodnu trgovinu, uključujući teretne, kontejnerske i putničke brodove. Ovakva klasifikacija dosljedno slijedi međunarodne standarde i omogućuje jasno razlikovanje skupina brodova prema veličini i operativnim karakteristikama.

Tablica se odnosi se na varijablu bruto tonaže broda koja predstavlja jednu od ključnih karakteristika u analizi pomorskih nezgoda jer direktno utječe na stabilnost broda, upravljivost, kapacitet tereta i ponašanje u različitim vremenskim uvjetima. Najveći broj zabilježen je u skupinama 1 i 4, što znači da su u nezgodama najčešće sudjelovali vrlo mali brodovi te srednje veliki trgovački. Kategorija 3 ima najmanji broj pojavljivanja, što može upućivati na njihovu manju zastupljenost u prometu ili manju učestalost nezgoda u usporedbi s drugim grupama. Tablica daje kvantitativnu sliku strukture podataka i dopunjuje klasifikaciju, omogućujući da se nezgode promatraju u odnosu na veličinu broda i da se identificiraju obrasci koji mogu biti relevantni za daljnje modele procjene rizika.

Tablica 23. Klasifikacija i broj pojavljivanja varijable *bruto tonaža* (*gt*)

Vrijednost	Raspon bruto tonaže (BT)	Opis klase broda	Broj pojavljivanja
1	0 – 99	Mala plovila, često obuhvaćaju rekreacijske i ribarske brodove male nosivosti	597

2	100 – 199,99	Brodovi srednje veličine, uglavnom manji trgovački ili servisni brodovi	183
3	200 – 499,99	Višefunkcionalni komercijalni brodovi, manji teretni brodovi	70
4	500 – 2999,99	Srednje do veći trgovački brodovi, uključujući RO-RO putničke i manje brodove za kontejnerski prijevoz	568
5	3000 i više	Veliki brodovi za međunarodni promet, teretni, kontejnerski i putnički brodovi	173

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Tablica 24. prikazuje klasifikaciju varijable *tip broda*, gdje je svaka numerička vrijednost pridružena određenoj vrsti broda. Naveden je engleski naziv i odgovarajući prijevod na hrvatski, čime je omogućeno jasno razumijevanje tipologije brodova uključenih u analizu. Prikazane su različite skupine, od teretnih, putničkih i ribarskih brodova, preko brodova za prijevoz tekućih tereta, tegljača i brodova za opskrbu odobalnih postrojenja, do rekreacijskih brodova i klase ostalo. Teretni brodovi imaju najveću zastupljenost s ukupno 490 slučajeva, što je u skladu s njihovom dominantnom ulogom u pomorskom prometu. Rekreacijski i servisni brodovi slijede s 184, odnosno 176 pojavljivanja, što upućuje na čestu prisutnost manjih brodova u situacijama rizika od nezgoda. Ribarski brodovi (167) i RO – RO putnički brodovi (104) također su značajno zastupljeni, dok su brodovi za prijevoz tekućih tereta ili kemikalija (116) te tegljači (94) prisutni u srednjem broju slučajeva. Putnički brodovi (77) i brodovi za opskrbu odobalnih postrojenja (58) čine manje zastupljene klase. Najrjeđe se pojavljuju kontejnerski brodovi (21), jaružala (25) i skupina ostalo (15). Ova raspodjela daje jasnu sliku strukture skupa podataka prema tipu broda.

Tablica 24. Klasifikacija i broj pojavljivanja varijable *tip broda* (*vessel_type*)

Vrijednost	Vrsta broda (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Bulk carrier*	Brod za rasuti teret	64
2	*Cargo ship*	Teretni brod	490
3	*Container ship*	Kontejnerski brod	21
4	*Dredger*	Jaružala	25
5	*Fishing vessel*	Ribarski brod	167
6	*RO–RO – passenger*	RO–RO putnički brod	104
7	*Tanker or chemical*	Brod za prijevoz tekućih tereta ili kemikalija	116
8	*Tug*	Tegljač	94
9	*Passenger*	Putnički brod	77
10	*Offshore vessel*	Brodovi za opskrbu odobalnih postrojenja	58
11	*Service ship*	Servisni brod	176
12	*Recreational craft*	Rekreacijska plovila	184
13	*Other*	Ostalo	15

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Tablica 25. prikazuje klasifikaciju varijable *duljina* broda, pri čemu su brodovi razvrstani u četiri jasno definirana raspona. Klase obuhvaćaju brodove do 50 m, od 50 do 100 m, od 100 do 200 m te one dulje od 200 m, što omogućuje jednostavno grupiranje brodova prema njihovim dimenzijama i olakšava analizu povezanosti duljine s pojavom pomorskih nezgoda. Najveću skupinu čine brodovi do 50 m duljine (1115 pojavljivanja), što ukazuje na dominaciju manjih brodova u analiziranom skupu podataka. Visoka zastupljenost manjih brodova može utjecati na odnose među varijablama koje modeli prepoznaju te na ukupnu uspješnost klasifikacije glavnih događaja. Stoga dobivene rezultate treba tumačiti u kontekstu strukture analizirane baze, uz oprez pri njihovu prenošenju na skupove s većim udjelom velikih brodova. Slijede brodovi duljine između 100 i 200 m (234), zatim oni između 50 i 100 m (138), dok su najrjeđe zastupljeni brodovi dulji od 200 m (104). Ova raspodjela potvrđuje da nezgode najčešće uključuju manje brodove.

Tablica 25. Klasifikacija i broj pojavljivanja varijable *duljina* broda (*length*)

Vrijednost	*length_(m)*	Objašnjenje	Broj pojavljivanja
1	(0,50]	duljina broda do 50 metara	1115
2	(50,100]	duljina broda između 50 i 100 m	138
3	(100,200]	duljina broda između 100 i 200 m	234
4	>200	duljina broda veća od 200 metara	104

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Tablica 26. prikazuje klasifikaciju i broj pojavljivanja varijable *širina* broda, kojom su brodovi razvrstani u tri raspona: do 20 m, između 20 i 30 m te više od 30 m. Ova podjela omogućuje pregledno grupiranje brodova prema njihovoj širini, što je važno za procjenu stabilnosti i manevarskih karakteristika broda. Tablica prikazuje raspodjelu brodova unutar navedenih kategorija. Najveći broj brodova nalazi se u prvoj skupini, sa širinom do 20 metara (1256 pojavljivanja), što potvrđuje dominaciju manjih brodova u analiziranom uzorku. Brodovi širine između 20 i 30 m (236) čine znatno manji udio, dok je grupa širine veće od 30 m najmanje zastupljena (99). Ovaj rezultat jasno ukazuje da su u nezgodama najčešće uključeni uži i manji brodovi.

Tablica 26. Klasifikacija i broj pojavljivanja varijable *širina* broda (*width)*

Vrijednost	*width_(m)*	Objašnjenje	Broj pojavljivanja
1	(0,20]	širina broda do 20 metara	1256
2	(20,30]	širina broda između 20 i 30 m	236
3	>30	širina broda veća od 30 metara	99

Izvor: izradio autor (EMSA, 2024; Li, 2023)

Varijabla *širina broda* zadržava se u analizi jer, unatoč očekivanoj korelaciji s duljinom, može pružiti dodatne informacije o stabilnosti, manevrabilnosti i manevarskim ograničenjima plovila, čime se omogućuje potpunije sagledavanje čimbenika rizika pomorskih nezgoda.

Tablica 27. prikazuje klasifikaciju i broj pojavljivanja varijable *vrsta štete*, pri čemu su nezgode razvrstane na materijalnu štetu, gubitak broda, manju štetu i slučajeve

bez zabilježene štete. Ovakva klasifikacija omogućuje jasnu procjenu ozbiljnosti posljedica pojedine nezgode. Najveći broj slučajeva odnosi se na nezgode bez štete (764), dok je manja šteta zabilježena u 321 slučaju. Materijalna šteta pojavljuje se 178 puta, a gubitak broda u 56 slučajeva. Značajan broj unosa (272) ima vrijednost *NA*, što upućuje na nedostajuće podatke u službenim zapisima.

Tablica 27. Klasifikacija i broj pojavljivanja varijable *vrsta štete* (*damage*)

Vrijednost	Opis (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*material damage*	materijalna šteta	178
2	*loss of ship*	gubitak broda	56
3	*minor damage*	manja šteta	321
4	*no damage*	bez štete	764
NA	*Unknown*	Nedostupno	272

Izvor: izradio autor (EMSA, 2024)

Tablica 28. prikazuje klasifikaciju i broj pojavljivanja varijable *onečišćenje*, koja obuhvaća tri osnovne vrijednosti: Da, Ne i Nije primjenjivo (engl. *Not Applicable – N/A*), uz dodatnu oznaku *NA* za nepoznate podatke. Ovakva klasifikacija omogućuje razlikovanje slučajeva u kojima je došlo do onečišćenja mora od situacija u kojima onečišćenje nije moglo nastati ili nije relevantno za vrstu nezgoda. U uvjerljivo najvećem broju slučajeva evidentirano je da nije došlo do onečišćenja (1175), dok je potvrđeno onečišćenje prisutno u 47 nezgoda. Kategorija *nije primjenjivo* obuhvaća 76 zapisa, uglavnom kod događaja gdje mogućnost onečišćenja ne postoji. Vrijednost *NA* zabilježena je u 293 slučaja, što ukazuje na nedostatne ili neunesene podatke.

Analiza pokazuje da je udio stvarnih onečišćenja relativno nizak, no svaki pojedinačni slučaj i dalje predstavlja značajan ekološki i sigurnosni rizik zbog osjetljivosti morskog okoliša i regulatornih obveza u njegovoj zaštiti.

Tablica 28. Klasifikacija i broj pojavljivanja varijable *onečišćenje* (*pollution*)

Vrijednost	***pollution***	Prijevod (hrvatski)	Broj pojavljivanja
1	Yes	Da	47
2	No	Ne	1175
3	N/A	Nije primjenjivo	76
NA	NA	Nepoznato	293

Izvor: izradio autor (EMSA, 2024)

Tablica 29. prikazuje klasifikaciju i broj pojavljivanja varijable *vrsta materijala broda*. Klase obuhvaćaju leguru aluminija, kompozitne materijale, laganu leguru, čelik, drvo i plastiku, što omogućuje jasno razlikovanje brodova prema materijalu gradnje i potencijalnim karakteristikama otpornosti i stabilnosti. Najveći broj brodova izrađen je od čelika (1141), dok je brodova od plastike 305, a od legura aluminija 71 brod. Lagana legura zastupljena je u svega 5 slučajeva, dok kompozitni materijali nisu zabilježeni. Brodovi od drva pojavljuju se u 34 slučaja, a 35 unosa ima vrijednost *NA*, što označava nedostatak podataka. Ova raspodjela pokazuje dominaciju čeličnih brodova u analiziranom skupu, dok su ostali materijali znatno rjeđe zastupljeni.

Tablica 29. Klasifikacija i broj pojavljivanja varijable *vrsta materijala broda* (*material*)

Vrijednost	Opis (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Aluminium alloy*	Legura aluminija	71
2	*composite materials*	Kompozitni materijali	0
3	*light alloy*	Lagana legura	5
4	*steel*	Čelik	1141
5	*wood*	Drvo	34
6	*plastic*	Plastika	305

Izvor: izradio autor prema (EMSA, 2024; Li, 2023)

Tablica 30. prikazuje klasifikaciju i broj pojavljivanja varijable *starost broda*, razvrstavajući brodove u pet dobnih skupina: do 5 godina, od 6 do 10 godina, od 11 do 15 godina, od 16 do 20 godina te preko 20 godina, uz dodatnu kategoriju *NA* za nepoznatu starost. Ova klasifikacija omogućuje pregled starosne strukture brodova i analizu potencijalne povezanosti starosti s učestalošću nezgoda. Najviše brodova pripada skupini starijih od 20 godina (497), dok su najmanje zastupljeni najmlađi brodovi do 5 godina (125). Brodovi starosti od 6 do 10 godina (213), 11 do 15 godina (199) i 16 do 20 godina (221) imaju srednju zastupljenost. Značajan broj unosa (336) označen je kao *NA*, što ukazuje na nedostatak podataka o starosti. Rezultati pokazuju da u analiziranom skupu dominiraju stariji brodovi, što može imati utjecaj na sigurnost i održavanje brodova.

Tablica 30. Klasifikacija i broj pojavljivanja varijable *starost broda* (*ship_age*)

Vrijednost	Starost broda (godine)	Objašnjenje	Broj pojavljivanja
1	[0,5]	do 5 godina	125
2	[6,10]	od 6 do 10 godina	213
3	[11,15]	od 11 do 15 godina	199
4	[16,20]	od 16 do 20 godina	221
5	>20	preko 20 godina	497
NA	NA	nepoznata starost broda	336

Izvor: izradio autor (Li, 2023)

Tablica 31. prikazuje klasifikaciju i broj pojavljivanja varijable *potonuće broda*, kojom se evidentira je li brod potonuo (*Da*) ili nije potonuo (*Ne*), uz dodatnu kategoriju *NA* za nedostajajuće ili nepoznate podatke. Ova podjela omogućuje pregled učestalosti potonuća brodova u analiziranom skupu podataka. Najveći broj brodova nije potonuo (1204), dok je potonuće potvrđeno u 57 slučajeva. Za 330 brodova podaci o potonuću su nedostajajući. Rezultati jasno prikazuju učestalost potonuća unutar skupa podataka, bez dodatnih interpretacija ili procjena uzroka nezgoda.

Tablica 31. Klasifikacija i broj pojavljivanja varijable *potonuće broda* (*vessel_sink)*

Vrijednost	Status potonuća (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	Yes	Da (brod je potonuo)	57
2	No	Ne (brod nije potonuo)	1204
NA	NA	Nedostajajuće / Nepoznato	330

Izvor: izradio autor (EMSA, 2024; Li, 2023)

Tablica 32. prikazuje klasifikaciju i broj pojavljivanja varijable *broj članova posade* razvrstavajući posade u pet kategorija: od 0 do 5, 6 do 10, 11 do 15, 16 do 20 i 21 ili više članova. Ova klasifikacija omogućuje sustavan pregled raspodjele posada po veličini broda i olakšava analizu učestalosti nezgoda u odnosu na broj članova posade. Najzastupljenija je prva kategorija, s posadama do 5 članova (594), dok su veće posade rjeđe zastupljene: od 6 do 10 članova (117), 11 do 15 članova (49), 16 do 20 članova (41) i 21 i više članova (83). Za značajan broj brodova (707) podaci o broju članova posade su nedostajajući. Podaci ukazuju na to da se većina nezgoda događa na manjim brodovima s manjom posadom, dok brodovi s većim brojem članova posade sudjeluju rjeđe, pri čemu je važno zadržati objektivan prikaz i ne izvlačiti zaključke o stručnosti posade ili uzrocima nezgoda.

Tablica 32. Klasifikacija i broj pojavljivanja varijable *broj članova posade* (*crew*)

Vrijednost	Broj članova posade	Broj pojavljivanja
1	Od 0 do 5	594
2	Od 6 do 10	117
3	Od 11 do 15	49
4	Od 16 do 20	41
5	21 i više	83
NA	Nepoznato	707

Izvor: izradio autor

Tablica 33. prikazuje klasifikaciju i broj pojavljivanja varijable *segment plovidbe* kojom su nezgode razvrstane prema fazi ili lokaciji plovidbe: dolazak, odlazak, tranzit, dvosmjerni plovni put, sidrenje ili pristajanje, preporučena ruta, zona pojačanog opreza, uz dodatnu kategoriju *Unknown* za nepoznate podatke. Ova podjela omogućuje sistematski pregled raspodjele nezgoda u različitim segmentima plovidbe i identifikaciju područja s većim rizikom. Najviše nezgoda zabilježeno je u kategorijama dolazak (440) i zona pojačanog opreza (403), dok ostali segmenti imaju manji broj pojavljivanja: odlazak (250), sidrenje ili pristajanje (141), preporučena ruta (36), tranzit (26) i dvosmjerni plovni put (14). Za 281 brod status segmenta plovidbe nije poznat (*NA*).

Pri interpretaciji rezultata važno je naglasiti da klase dolazak i odlazak predstavljaju opće faze plovidbe koje se mogu pojaviti u svim područjima plovidbe, dok ostale klase definiraju specifična područja ili uvjete navigacije. Zbog toga je argumentirano razmatranje spajanja dolaska i odlaska u jednu kategoriju u analizi moguće, no za deskriptivni prikaz skupa podataka ostaju zasebno prikazane radi jasnoće i potpunosti evidencije. Također uplovljenje predstavlja fazu plovidbe s povećanim rizikom od nezgode zbog kombinacije postepenog smanjivanja brzine, smanjenog manevarskog prostora, visoke gustoće prometa, hidrodinamičkih interakcija s obalom i lučkom infrastrukturom te potrebe za preciznom kontrolom brzine i kursa. Nasuprot tome, isplovljenje se odvija u uvjetima progresivnog povećanja raspoloživog prostora i smanjenja operativnih ograničenja, stoga su ove dvije faze plovidbe razdvojene kao zasebne klase. Na ovaj način se osigurava objektivna prezentacija podataka bez subjektivnog ocjenjivanja rizika ili uzroka nezgoda.

Tablica 33. Klasifikacija i broj pojavljivanja varijable *segment plovidbe* (*voyage_segment*)

Vrijednost	Opis (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Arrival*	Dolazak	440
2	*Departure*	Odlazak	250
3	*Transit*	Tranzit	26
4	*Two–way route*	Dvosmjerni plovni put	14
5	*Anchor or alongside*	Sidrenje ili pristajanje	141
6	*Recommended route*	Preporučena ruta	36
7	*Precautionary area*	Zona pojačanog opreza	403
NA	*Unknown*	Nepoznato	281

Izvor: izradio autor prema (EMSA, 2024)

Tablica 34. prikazuje jasnu klasifikaciju i broj pojavljivanja varijable *peljar na brodu* s odgovarajućim prijevodima. Uključene su klase za neprimjenjive i nepoznate podatke, što osigurava potpunost klasifikacije. Brojevi jasno pokazuju raspodjelu slučajeva s peljarom, bez peljara te onih za koje podatak nije primjenjiv ili poznat. Ovakva tablica omogućuje uvid u učestalost i zastupljenost pojedinih kategorija, što je važno za daljnju statističku obradu ili modeliranje. Zanimljivo je da postoji i značajan broj nedostajućih vrijednosti (*NA*), koje nose oznaku *Nepoznato* i pojavljuju se 250 puta. Ova distribucija ukazuje na činjenicu da u velikom broju slučajeva status peljara ili nije relevantan za kontekst (segment plovidbe), ili podaci o njegovoj prisutnosti su nedostajajući.

Tablica 34. Klasifikacija i broj pojavljivanja varijable *peljar na brodu* (*pilot*)

Vrijednost	Opis (engleski)	Prijevod (hrvatski)	Broj pojavljivanja
1	*Pilot on board*	Peljar na brodu	480
2	*Pilot not on board*	Peljar nije na brodu	320
3	*N/A*	Nije primjenljivo	541
NA	*Unknown*	Nepoznato	250

Izvor: izradio autor (EMSA, 2024)

Tablica 35. prikazuje klasifikaciju i broj pojavljivanja varijable *ukupno ozljeda na brodu*, koja označava ukupan broj osoba ozlijeđenih u pojedinoj nezgodi. Varijabla je podijeljena u četiri klase: vrijednost 1 označava da nije bilo ozlijeđenih, vrijednost 2 označava jednu ozljedu, vrijednost 3 dvije ozljede, a vrijednost 4 tri ili više ozljeda. Takva klasifikacija omogućuje sustavno evidentiranje ozbiljnosti nezgode i olakšava analize i modeliranje podataka, zadržavajući ključnu informaciju o posljedicama događaja. Većina zapisa (1387) pripada kategoriji 1, što znači da u tim slučajevima nije bilo ozlijeđenih. Klase s jednom, dvije i tri ili više ozljeda pojavljuju se znatno rjeđe, sa ukupno 51 slučajem. Broj nedostajućih podataka (*NA*) iznosi 153, što predstavlja manji, ali značajan udio koji treba uzeti u obzir pri obradi i analizi podataka. Podaci iz tablice ukazuju da većina pomorskih nezgoda ne rezultira ozljedama, dok su ozbiljniji događaji s višestrukim ozljedama rijetki. Ova analiza pruža uvid u strukturu rizika i važna je za procjenu i planiranje sigurnosnih mjera na brodovima.

Tablica 35. Klasifikacija i broj pojavljivanja varijable *ukupno ozljeda na brodu* (*injuries_total*)

Vrijednost	Broj ozljeda	Broj pojavljivanja
1	0	1387
2	1	23
3	2	5
4	3 i više	23
NA	Nepoznato	153

Izvor: izradio autor (EMSA, 2024)

U provedenim postupcima učitavanja podataka, transformacije varijabli, imputacije nedostajućih vrijednosti te izrade deskriptivne statistike korišteni su ključni paketi i funkcije programskog jezika R. Paketi *readxl* i *openxlsx* služile su za učitavanje izvornih Excel datoteka te generiranje i uređivanje Excel izvještaja. U tu svrhu primijenjena je funkcija *read_excel* za uvoz podataka, dok je petlja *for* korištena za dosljednu konverziju varijabli u tekstualni oblik kada nisu bile definirane kao faktorske ili tekstualne.

Za analizu i opis podatkovnih struktura korištene su korisnički definirane procedure *get_mode_info* (izračun modusa i njegove frekvencije), *describe_factors_chars* (detaljna deskripcija faktorskih i tekstualnih varijabli), *frequency_tables* (generiranje frekvencijskih tablica) te *sorted_frequencies* (prikaz sortiranih frekvencija). Funkcija *prepare_for_description* služila je za pripremu podataka u format pogodan za strukturirano prikazivanje, dok je imputacija nedostajućih vrijednosti provedena unutar svake razine ciljne varijable *main_event*.

Za izradu konačnih Excel izvještaja korištene su funkcije *createWorkbook*, *addWorksheet*, *writeData* i *saveWorkbook*, čime je osiguran potpun, reproducibilan i pregledan zapis svih obrađenih rezultata.

Ukupno je analizirano 1591 evidencija i 29 varijabli, pri čemu je varijabla *glavni događaj* određena kao ciljna, dok ostale varijable služe kao ulazne značajke. Privremenim numeričkim kodiranjem olakšana je organizacija podataka i provedba statističkih provjera, uz očuvanje svih informacija i osiguranje pravilnog tretmana kategorijskih značajki.

U narednom potpoglavlju posebna pažnja bila je posvećena transformaciji numeričkih vrijednosti u deskriptivne oznake kod kategorijskih varijabli, kako bi se spriječilo pogrešno tumačenje redoslijeda i očuvala semantička vrijednost podataka.

3.2. Transformacija varijabli i imputacija nedostajućih vrijednosti

U procesu pripreme podataka za klasifikaciju pomorskih nezgoda provedena je kategorizacija odabranih numeričkih varijabli, standardizacija numeričkih varijabli, ujednačavanje kategorijskih oznaka te nadomještanje nedostajućih vrijednosti najčešćom vrijednošću pojedine varijable unutar odgovarajuće klase glavnog događaja. Ovim

korakom osigurana je ispravna interpretacija podataka, stabilnost modela i valjanost rezultata.

Tablici B1. – Prilog B daje sustavan i metodološki utemeljen pregled svih 29 varijabli korištenih u empirijskoj analizi pomorskih nezgoda, s ciljem jasnog razgraničenja njihove stvarne mjerne prirode i načina obrade u statističkoj analizi. Iako su u izvornom skupu podataka gotovo sve varijable tehnički pohranjene kao faktori ili tekstualne varijable, takva informatička reprezentacija sama po sebi nije dovoljna za ispravan odabir statističkih metoda. Naime, način pohrane podataka ne mora odgovarati njihovoj stvarnoj statističkoj i metodološkoj naravi, što može dovesti do pogrešnog izbora mjera povezanosti i, posljedično, do pogrešne interpretacije rezultata. Upravo iz tog razloga bilo je nužno provesti dodatnu i preciznu klasifikaciju varijabli na ordinalne i nominalne, u skladu s ciljevima istraživanja i primijenjenim korelacijskim metodama, kako bi se omogućila učinkovita implementacija modela potpornih vektora, slučajnih šuma i neuronskih mreža za klasifikaciju nezgoda.

Osnovna svrha ove klasifikacije jest osigurati metodološku ispravnost korelacijske analize koja se provodi u narednoj fazi. Različite mjerne razine zahtijevaju primjenu različitih statističkih koeficijenata, budući da svaka mjera povezanosti počiva na specifičnim pretpostavkama o prirodi podataka. Primjena jedinstvenog korelacijskog koeficijenta na sve varijable, bez obzira na njihov tip, dovela bi do metodološki neopravdanih postupaka i interpretativno nepouzdanih zaključaka. Stoga je bilo nužno jasno razdvojiti varijable koje posjeduju prirodan poredak kategorija od onih kod kojih takav poredak ne postoji.

Tablica jasno diferencira dvije osnovne skupine varijabli. Prvu skupinu čine nominalne varijable, koje nemaju prirodni redoslijed među kategorijama i kod kojih su sve razine međusobno ravnopravne. Drugu skupinu čine ordinalne varijable, kod kojih postoji smislen i logički utemeljen poredak razina, ali ne i pretpostavka jednakih razmaka među njima. Ovakva diferencijacija predstavlja temeljni preduvjet za statistički ispravno mjerenje povezanosti između varijabli te čini polazište za daljnju analitičku obradu podataka.

Ordinalne varijable, kao što su *ozbiljnost nezgode*, *godina prijave*, *mjesec*, opisuju pojave kod kojih je hijerarhijski odnos između kategorija jasno definiran. Slična logika vrijedi i za varijable koje opisuju okolišne uvjete pomorske nezgode, poput *prirodna svjetlost*, *stanje mora*, *vidljivost*, *jačina vjetra*, gdje vrijednosti predstavljaju rastući intenzitet ili stupanj izraženosti promatranog fenomena. Njihova klasifikacija kao ordinalnih varijabli omogućuje zadržavanje semantičkog značenja podataka, uz istodobnu mogućnost kvantitativne analize.

U istu skupinu ubrajaju se i varijable koje opisuju tehničke karakteristike broda i posljedice nezgode, kao što su *bruto tonaža*, *duljina*, *širina*, *vrsta štete*, *onečišćenje*, *starost broda*, *broj članova posade*, *ukupno ozljeda na brodu*. Kategorije *bruto tonaže*, *duljine* i *širine* odgovaraju uzlaznim rasponima vrijednosti te se prema sadržaju mogu smatrati ordinalnima. Pri treniranju klasifikacijskih modela u programskom okruženju R

kodirane su kao neuređeni faktori, odnosno kategorijske varijable s definiranim razinama bez izričito zadanog poretka. Za korelacijsku analizu njihove su kategorijske oznake zasebno prikazane brojčanim redoslijedom raspona.

S druge strane, nominalne varijable, poput *glavni događaj*, *baza podataka*, *država izvjetiteljica*, *vrijeme*, *lokacija nezgode*, *područje nezgode*, *zastava broda*, *kategorija broda*, *tip broda*, *materijal, potonuće broda, segment putovanja*, *peljar*, predstavljaju čisto kvalitativna obilježja bez prirodnog poretka među kategorijama. Tretiranje brojčanih oznaka ovih nominalnih varijabli kao numeričkih veličina nametnulo bi poredak i udaljenosti među kategorijama koji nemaju sadržajno opravdanje. Stoga ove varijable u programskom okruženju R ostaju neuređeni faktori, odnosno kategorijske varijable s definiranim razinama bez zadanog poretka. Njihova međusobna povezanost, kao i povezanost s ordinalnim varijablama, procjenjuje se pomoću Cramérovog V koeficijenta, prikladnog za analizu povezanosti kategorijskih podataka.

Time tablica ne predstavlja samo deskriptivni prikaz korištenih varijabli, već čini ključni metodološki element istraživanja. Ona osigurava transparentnost analitičkog postupka, omogućuje ponovljivost rezultata te jasno opravdava izbor statističkih metoda korištenih u korelacijskoj analizi. Na taj način postavlja čvrstu metodološku osnovu za interpretaciju dobivenih rezultata i osigurava njihovu znanstvenu valjanost i pouzdanost.

Metodološka dopuna prikazana u Tablici B1. – Prilog B osigurava pouzdanu klasifikaciju i kvantitativnu procjenu utjecaja pojedinih varijabli na predviđene ishode, uz očuvanje semantičke strukture podataka.

U izvornom skupu od 1591 zapisa nedostajalo je 834 vrijednosti varijable *vrijeme* (52,4 %), 724 vrijednosti varijable *vidljivost* (45,5 %), 707 vrijednosti varijable *broj članova posade* (44,4 %) te po 657 vrijednosti varijabli *stanje mora* i *jačina vjetra* (41,3 %). Takav opseg nedostajanja treba uzeti u obzir pri tumačenju povezanosti i rezultata modeliranja. Za rješavanje nedostajućih vrijednosti primijenjena je imputacija unutar svake klase ciljne varijable, odnosno razine rizika. Nedostajući podaci zamijenjeni su modalnom vrijednošću unutar pripadajuće klase, čime je očuvana originalna struktura distribucije i spriječeno unošenje pristranosti u modele. Varijable koje su bile potpuno popunjene, poput *država izvjestiteljica*, *područje nezgode* i *zastave broda*, nisu zahtijevale dodatnu obradu.

Pripremljeni skup podataka predstavljao je preduvjet za provođenje korelacijske analize, koja je omogućila prepoznavanje odnosa među ulaznim značajkama i identificiranje varijabli s potencijalno najsnažnijim utjecajem na razinu rizika. Korelacijska analiza provedena je kako bi se utvrdilo kako su numeričke, ordinalne i nominalne varijable međusobno povezane te kakve obrasce odnosa pokazuju unutar skupa podataka.

Završna faza ove cjeline potvrđuje da su provedene transformacije podataka i postupci imputacije stvorili čvrste metodološke temelje za pouzdanu korelacijsku analizu te za precizniju primjenu naprednih klasifikacijskih modela u daljnjim poglavljima. Poseban naglasak stavljen je na ispravnoj pripremi varijabli različitih mjernih razina kako bi se izbjegle metodološke pogreške u interpretaciji povezanosti među varijablama.

U ovoj fazi priprema podataka u programskom okruženju R obuhvatila je sustavno razgraničenje i obradu ordinalnih i nominalnih varijabli radi njihove pravilne statističke interpretacije i upotrebljivosti u korelacijskoj analizi. Izvorni skup podataka učitan je pomoću funkcije *read_excel()* iz paketa *readxl*, čime je osigurano precizno učitavanje podataka i očuvanje njihove izvorne strukture.
Varijable s prirodnim redoslijedom kategorija razlikovane su prema njihovu sadržaju, dok je njihov način kodiranja ovisio o analitičkom postupku. Pri treniranju modela varijable *bruto tonaža*, *duljina* i *širina* korištene su kao neuređeni faktori. Radi provođenja korelacijske analize, ordinalne varijable su privremeno transformirane u numerički oblik, isključivo u tehničke svrhe izračuna Pearsonovog i Spearmanovog koeficijenta korelacije. Time je omogućena kvantitativna procjena jačine i smjera povezanosti, uz zadržavanje logike ordinalnog poretka i bez pretpostavke kontinuiranosti podataka. Nominalne varijable nisu numerički transformirane, već je njihova međusobna povezanost, kao i povezanost s ordinalnim varijablama, procjenjivana pomoću Cramérovog V koeficijenta, koji je prikladan za analizu asocijacije među kategorijskim podacima.

Korelacijska analiza provedena je odvojeno za skup ordinalnih varijabli, za skup nominalnih varijabli te za kombinacije ordinalnih i nominalnih varijabli. Ovakav pristup omogućio je razumijevanje obrazaca povezanosti unutar skupa podataka, bez metodološkog utjecaja tipa varijable na izračun korelacijskih mjera.
U ovoj fazi provedena je i imputacija nedostajućih vrijednosti unutar pojedinih klasa ciljne varijable *glavni događaj*, čime je očuvana struktura distribucije podataka i smanjen rizik uvođenja pristranosti u daljnje analitičke modele. Za masovnu obradu varijabli korištene su funkcije *lapply()* i iterativne *for()* petlje, dok je paket *openxlsx* korišten za generiranje Excel izvještaja i pohranu rezultata korelacijske analize.

Cjelokupni pristup osigurao je da podaci budu standardizirani, konzistentni i metodološki ispravno pripremljeni za korelacijske analize te za daljnju primjenu naprednih klasifikacijskih modela u procjeni rizika pomorskih nezgoda. Na tako strukturiranim podacima primijenjena su tri modela strojnog učenja, i to: model potpornih vektora, model slučajnih šuma i model neuronskih mreža. Sva tri modela korištena su isključivo za klasifikaciju pomorskih nezgoda prema tipu glavnog događaja, pri čemu je svaki model omogućio drugačiju perspektivu u pogledu interpretacije ulaznih varijabli, nelinearnih odnosa i robusnosti predikcija. Nakon provedene klasifikacije tipova glavnih događaja, isti modeli dodatno su primijenjeni za izračun vjerojatnosti za sedam različitih posljedičnih događaja definiranih kao glavni događaji, čime je omogućena kvantifikacija vjerovatnosti po pojedinim ishodima nezgoda u obalnoj plovidbi.

3.3. Korelacijska analiza promatranih čimbenika

Završna faza pripreme podataka omogućila je da skup podataka bude standardiziran, konzistentan i metodološki ispravno obrađen, čime su stvoreni čvrsti temelji za provođenje korelacijske analize. Cilj ove analize je prepoznati međusobne odnose među

ulaznim varijablama, identificirati one s potencijalno najznačajnijim utjecajem na razinu rizika te osigurati temelj za optimiziranu selekciju prediktora u modelima strojnog učenja.

Korelacijska analiza predstavlja ključni korak u razumijevanju strukture podataka jer omogućava kvantifikaciju snage i smjera povezanosti između varijabli, otkrivanje potencijalne redundantnosti i multikolinearnosti te identifikaciju obrazaca koji mogu utjecati na interpretaciju rezultata modela. Posebno je važno da se metode odaberu prema tipu varijabli: ordinalne i numeričke varijable zahtijevaju drugačiji pristup od nominalnih varijabli, dok kombinacija nominalnih i ordinalnih/numeričkih varijabli zahtijeva primjenu koeficijenata koji su prilagođeni mješovitim tipovima podataka.

Za procjenu povezanosti između numeričkih i ordinalnih varijabli primijenjeni su Pearsonov i Spearmanov koeficijent korelacije, čime se omogućuje ocjena linearnih i monotonih odnosa među promatranim čimbenicima. Za analizu povezanosti među nominalnim varijablama, kao i za relacije između nominalnih i ordinalnih/numeričkih varijabli, korišten je Cramérov V koeficijent, koji je metodološki prikladan za kategorizirane podatke bez prirodnog poretka.

Za analizu međusobne povezanosti numeričkih varijabli primijenjen je pristup odabira metode prema vrsti podataka, što omogućava preciznu i robusnu procjenu veza među varijablama. Spearmanov rang-koeficijent korelacije (ρ) koristi se za ordinalne ili neparametarske varijable te ocjenjuje snagu i smjer monotonih odnosa bez pretpostavke linearnosti, temeljeći se na rangovima podataka (Formula (24)). Transformacijom podataka u rangove uklanja se osjetljivost na odstupanja i ekstremne vrijednosti, što je posebno korisno u analizama pomorskih baza podataka s kategoriziranim ili ordinalnim varijablama (Mukaka, 2012; Schober et al., 2018).

$$\rho = 1 - \frac{6\sum d_1^2}{n(n^2-1)} \tag{24}$$

gdje je d_i razlika između rangova za svaki par opažanja, a n broj opažanja.

Pearsonov koeficijent korelacije (r) primjenjuje se također za numeričke varijable s normalnom distribucijom, kvantificirajući linearnu povezanost između kontinuiranih varijabli (Formule (25) i (26)). Vrijednosti koeficijenta pružaju uvid u intenzitet i smjer linearne povezanosti, pri čemu se pretpostavlja normalna distribucija, linearna ovisnost, homogenost varijanci i odsustvo ekstremnih vrijednosti. Korištenje oba koeficijenta omogućuje identifikaciju i linearnih i monotonih nelinearnih veza, čime se povećava pouzdanost odabira prediktora za modele procjene rizika.

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \cdot \sqrt{\sum(y_i-\bar{y})^2}}} \tag{25}$$

gdje su:

- r – Pearsonov koeficijent korelacije,
- x_i, y_i – pojedinačne vrijednosti varijabli,
- $\bar{x}$, $\bar{y}$ – aritmetičke sredine varijabli x i y,
- brojnik predstavlja kovarijancu između varijabli,
- nazivnik je produkt njihovih standardnih devijacija.

Alternativno, kada se radi s gotovim statističkim vrijednostima, Pearsonov koeficijent može se izraziti i kao u Formuli (26):

$$r = \frac{cov(X,Y)}{\sigma X \cdot \sigma Y} \tag{26}$$

gdje su:

- *cov(X,Y)* – kovarijanca između varijabli *X* i *Y*,
- *σX, σY* – standardne devijacije varijabli *X* i *Y*.

Za kategorijske varijable primijenjen je χ^2 (engl. *Chi – Square Test of Independence*), test nezavisnosti (Formula (27)), koji ispituje statističku značajnost povezanosti između dviju nominalnih varijabli uspoređujući opažene i očekivane frekvencije, pri čemu se očekivane frekvencije računaju prema Formuli (28), *p* – manje od 0,05 ukazuje na statistički značajnu povezanost (Kassambara, 2018).

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{27}$$

gdje su:

- χ^2 – vrijednost testne statistike,
- O_{ij} – opažena frekvencija u ćeliji *i,j* kontingencijske tablice,
- E_{ij} – očekivana frekvencija u ćeliji *i,j*, računata po Formuli (28).

$$E_{ij} = \frac{(R_i \cdot C_j)}{N} \tag{28}$$

- R_i – zbroj frekvencija u *i*-tom redu,
- C_j – zbroj frekvencija u *j*-toj koloni,
- *N* – ukupan broj slučajeva.

Napomena:
χ^2 – vrijednost hi-kvadrat testa (χ^2 test), se koristi za procjenu statističke značajnosti razlika između očekivanih i opaženih frekvencija u kategorijskim podacima. Veće vrijednosti χ^2 ukazuju na veće odstupanje od očekivanih vrijednosti, pri čemu se uzima u obzir i pripadajuća razina značajnosti (*p* – vrijednost) za donošenje zaključaka.

Kako bi se kvantificirala jačina povezanosti među nominalnim varijablama, korišten je Cramérov V koeficijent (V) (Formula (29)). Ovaj standardizirani koeficijent temeljen na χ^2 vrijednosti omogućava interpretaciju snage povezanosti neovisno o veličini uzorka i broju kategorija. Vrijednosti bliže 0 ukazuju na slabu povezanost, a bliže 1 na snažnu povezanost, što ga čini prikladnim za preliminarnu analizu kategorijskih prediktora u kontekstu procjene rizika.

$$V = \sqrt{\frac{\chi^2}{n \cdot (k-1)}} \tag{29}$$

gdje su:

- χ^2 – vrijednost hi-kvadrat testa,
- *n* – ukupan broj promatranih slučajeva,
- *k* – manji broj redaka ili stupaca u kontingencijskoj tablici (ovisno o dimenzijama).

Cramérov V koeficijent primjenjiv je na sve nominalne varijable jer ne pretpostavlja linearnost, što ga čini pogodnim za preliminarnu analizu u svrhu identifikacije značajnih veza između kategorijskih prediktora i ciljne varijable. U kontekstu strojnog učenja, visoke vrijednosti Cramérovog V mogu ukazivati na redundanciju među prediktorima i potencijalnu multikolinearnost u modelima poput logističke regresije ili neuronskih mreža, dok niske vrijednosti sugeriraju varijable niske prediktivne moći koje mogu povećati šum u modelu (Akoglu, 2018; Kuhn & Johnson, 2013). Osim procjene povezanosti, Cramérov V koristi se u odabiru značajki, eliminaciji irelevantnih prediktora i smanjenju dimenzionalnosti skupa podataka, čime se optimiziraju prediktivni modeli (Brownlee, 2020).

Vrijednosti Cramérovog V koeficijenta kreću se u rasponu od 0 do 1, pri čemu:

- vrijednosti bliže 0 ukazuju na slabu povezanost između varijabli,
- vrijednosti bliže 1 ukazuju na snažnu povezanost,
- vrijednost 0 označava potpunu nezavisnost, dok vrijednost 1 označava savršenu povezanost, ali samo u slučajevima kad je broj redaka jednak broju stupaca.

Slojeviti pristup, kombinacija Spearmanove i Pearsonove korelacije za numeričke varijable te χ^2 testa i Cramérovog V za kategorijske varijable, omogućava sveobuhvatnu, robustnu i transparentnu analizu međusobnih odnosa među varijablama, što je ključno za interpretaciju rezultata i konstrukciju preciznih modela strojnog učenja u domenu pomorskih rizika.

Korelacijska analiza varijabli provedena je korištenjem paketa *ggcorrplot* i *corrplot* za vizualizaciju matrica korelacija numeričkih varijabli za kategorijske odnose, čime je olakšana identifikacija kolinearnosti i izbor relevantnih prediktora. Podaci su učitani paketom *readxl*, numerički rezultati zapisani paketom *openxlsx*, dok su za izračun Cramérova V korišteni *DescTools*. Vizualizacija je provedena paketom *ggplot2*, transformacija matrica paketom *reshape2*, a priprema najsnažnijih 10 korelacija i manipulacija podacima paketom *dplyr*. Sačuvani numerički podaci osigurali su temeljnu podlogu za daljnju podjelu podataka i pouzdano modeliranje pomorskih rizika (RStudio Team, 2024).

Glavni cilj korelacijske analize bio je dvostruk: pružiti dublje razumijevanje odnosa među varijablama u kontekstu sigurnosti obalne plovidbe i identificirati varijable s visokom međusobnom povezanosti, čime se dobiva uvid u moguće međuzavisnosti koje mogu utjecati na interpretaciju i performanse modela (Kuhn & Johnson, 2013; Brownlee, 2020).

3.4. Podjela podataka na trenirajući, validacijski i testni skup

U ovoj fazi istraživanja provedena je podjela ukupnog skupa podataka na tri nezavisna podskupa: trenirajući skup koji obuhvaća 70% podataka (1115 slogova), validacijski skup s 15% podataka (240 slogova) i testni skup također s 15% podataka (236 slogova), što ukupno iznosi 1591 slog. Podjela je izvršena metodom stratificiranog slučajnog uzorkovanja, čime je očuvana distribucija ciljne varijable *glavni događaj*. Kontrola

nasumičnosti provedena je fiksiranjem generatora slučajnih brojeva putem funkcije *set.seed*, što omogućuje potpunu reproduktivnost rezultata. Takva organizacija podataka ključna je za objektivnu procjenu sposobnosti modela da generalizira na nove primjere i smanjenje rizika prenaučenosti (Kuhn & Johnson, 2013).

Trenirajući skup koristi se za učenje parametara modela, validacijski skup omogućuje podešavanje hiperparametara i usporedbu različitih konfiguracija modela bez utjecaja na konačnu evaluaciju, dok se testni skup koristi isključivo za završnu procjenu kvalitete modela, osiguravajući nepristranu i pouzdanu mjeru njegove učinkovitosti. Podjela je izvedena uz pomoć funkcije *createDataPartition* iz R paketa *caret*, koja osigurava stratificiranu raspodjelu podataka te održava omjer klasa ciljne varijable gotovo identičnim u svim podskupovima. Time se sprječava iskrivljena evaluacija i smanjuje rizik favoriziranja dominantnih klasa.
Postupak započinje fiksiranjem generatora slučajnih brojeva (*set.seed(123)*) kako bi svaki ponovni izračun rezultirao istom podjelom podataka. Prvo se izdvaja 70% podataka za treniranje, dok se preostalih 30% ravnomjerno raspodjeljuje na validacijski i testni skup.

Validacijski skup koristi se za izbor optimalnih hiperparametara, a testni ostaje potpuno izdvojen sve do završne evaluacije. Nakon odabira optimalnih hiperparametara, model se trenira na proširenom skupu *train_valid_data*, koji uključuje trenirajući i validacijski skup zajedno, čime se maksimizira količina podataka dostupna za konačnu izgradnju modela, ali tek nakon što je proces validacije proveden na strogo odvojenom skupu.

Podjela podataka na trenirajući, validacijski i testni skup predstavlja ključnu fazu u izgradnji robusnih i reproducibilnih modela strojnog učenja, omogućujući neovisnu evaluaciju i precizno procjenjivanje sposobnosti modela da generalizira na nove podatke.

3.5. Metrike procjene performansi modela i predikcije ishoda

Evaluacija modela strojnog učenja temelji se na preciznoj procjeni njihove sposobnosti predviđanja ishoda. Svi modeli klasifikacije rizika su procijenjeni na testnom skupu podataka koristeći proširenu matricu zabune dimenzija 7×7, što odgovara broju kategorija ciljne varijable *glavni događaj*.

Matrica zabune (Tablica 36.) predstavlja osnovni analitički alat za ocjenu performansi klasifikacijskih modela, pri čemu kvadratnom strukturom prikazuje koliko je instanci iz stvarnih klasa ispravno ili nepravilno klasificirano u predviđene klase. Ovako višedimenzionalno sagledavanje omogućuje detaljan uvid u ponašanje modela i sprječava oslanjanje isključivo na jednu metriku, poput ukupne točnosti, koja može prikriti nedostatke u klasifikaciji pojedinih klasa.

Tablica 36. Matrica zabune

Elementi Matrice zabune	**Predviđeno: Pozitivno (engl. *Predicted Positive* – PP)**	**Predviđeno: Negativno (engl. *Predicted Negative* – PN)**
Stvarno: Pozitivno (engl. *Actual Positive* – AP)	Istiniti pozitivni (engl. *True Positive* – TP)	Lažno negativni (engl. *False Negative* – FN)

Stvarno: Negativno (engl. *Actual Negative* – AN)	Lažno pozitivni (engl. *False Positive* – FP)	Istiniti negativni (engl. *True Negative* – TN)

Izvor: izradio autor

Ciljna varijabla *glavni događaj* podijeljena je u sedam kategorija koje predstavljaju različite vrste glavnih događaja u sigurnosti obalne plovidbe. Proširena matrica zabune dimenzija 7×7 koristi retke za stvarne klase, a stupce za predviđene. Svaka ćelija matrice prikazuje broj primjera koji stvarno pripadaju određenoj klasi te koliko ih je model predvidio u svaku od sedam klasa. Dijagonalne vrijednosti ukazuju na ispravna predviđanja, dok visoke vrijednosti izvan dijagonale ukazuju na zabunu modela između određenih klasa. Takva analiza omogućuje identifikaciju klasa koje model najbolje predviđa, kao i problema u klasifikaciji, čime se stvara osnova za izračun različitih metrika kvalitete modela.

U Tablici 37., u prvom stupcu je redni broj metrike, u drugom stupcu je naziv metrike na hrvatskom jeziku uz engleski ekvivalent u zagradi, u trećem stupcu je opis metrike koji objašnjava što metrike mjere i čemu služe u evaluaciji klasifikacijskih modela, a u četvrtom stupcu je formula za izračun metrike koristeći označene pojmove: točno pozitivni (engl. *true positives* – TP), točno negativni (engl. *true negatives* – TN), netočno pozitivni (engl. *false positives* – FP) i netočno negativni (engl. *false negatives* – FN).

Na temelju elemenata matrice zabune izračunato je četrnaest metrika performansi među kojima su: točnost, preciznost, osjetljivost, F1-mjera, specifičnost, negativna prediktivna vrijednost, stopa lažno pozitivnih, lažno negativnih predikcija, omjer pozitivne vjerojatnosti i druge (James et al., 2021). Ove metrike daju numerički pokazatelj koliko dobro model klasificira slučajeve u svaku klasu. Za sažeti prikaz osjetljivosti, preciznosti i F1-mjere u višeklasnom problemu izračunava se aritmetička sredina odgovarajućih vrijednosti za svih sedam klasa, pri čemu svaka klasa ima jednaku težinu (makroprosjek). Ukupna točnost zasebno se računa kao omjer broja točno klasificiranih i ukupnog broja zapisa.

Tablica 37. Korištene metrike za ocjenu klasifikacijskih modela

Rb.	Naziv metrike	Opis	Formula
1	Točnost (engl. *Accuracy* – ACC)	Udio točno klasificiranih slučajeva u odnosu na ukupni broj instanci.	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$ (30)
2	Stopa pogreške (engl. *Error Rate*)	Ukupni udio netočno klasificiranih instanci u odnosu na sve instance.	$ErrorRate = \frac{FP+FN}{TP+TN+FP+FN}$ (31)
3	Osjetljivost (engl. *Sensitivity* ili *True Positive Rate* – TPR)	Sposobnost modela da prepozna instancu koja pripada pozitivnoj klasi.	$TPR = \frac{TP}{TP + FN}$ (32)
4	Specifičnost (engl. *Specificity* ili *True Negative Rate* – TNR)	Sposobnost modela da ispravno prepozna instancu koja ne pripada pozitivnoj klasi.	$TNR = \frac{TN}{TN + FP}$ (33)

5	Preciznost (engl. *Precision* ili *Positive Predictive Value* – PPV)	Udio ispravno klasificiranih pozitivnih instanci u odnosu na sve koje su klasificirane kao pozitivne.	$PPV = \frac{TP}{TP+FP}$ (34)
6	Negativna prediktivna vrijednost (engl. *Negative Predictive Value* – NPV)	Udio ispravno klasificiranih negativnih instanci u odnosu na sve koje su klasificirane kao negativne.	$NPV = \frac{TN}{TN+\mathrm{FN}}$ (35)
7	F1-mjera (engl. *F1 Score*)	Harmonijska sredina između preciznosti i osjetljivosti; koristi se kada je ravnoteža između FP i FN bitna.	$\mathrm{F1} = 2 \times \frac{PPV \times \mathrm{TRP}}{PPV+\mathrm{TPR}}$ (36)
8	Stopa lažno pozitivnih predikcija (engl. *False Discovery Rate* – FDR)	Udio instanci koje su pogrešno klasificirane kao pozitivne u odnosu na sve predviđene pozitivne.	$\mathrm{FDR} = \frac{FP}{TP+FP}$ (37)
9	Stopa lažno negativnih predikcija (engl. *False Omission Rate* – FOR)	Udio instanci koje su pogrešno klasificirane kao negativne u odnosu na sve predviđene negativne.	$\mathrm{FOR} = \frac{FN}{TN+FN}$ (38)
10	Stopa lažno pozitivnih slučajeva (engl. *False Positive Rate* – FPR)	Udio negativnih instanci koje su pogrešno klasificirane kao pozitivne.	$\mathrm{FPR} = \frac{FP}{FP+TN}$ (39)
11	Stopa lažno negativnih slučajeva (engl. *False Negative Rate* – FNR)	Udio pozitivnih instanci koje su pogrešno klasificirane kao negativne.	$\mathrm{FNR} = \frac{FN}{TP+FN}$ (40)
12	Omjer pozitivne vjerojatnosti (engl. *Positive Likelihood Ratio* – LR^+)	Koliko puta je vjerojatnije da će model ispravno klasificirati pozitivni slučaj u odnosu na pogrešnu klasifikaciju negativnog.	$\mathrm{LR}^+ = \frac{TPR}{FPR}$ (41)
13	Omjer negativne vjerojatnosti (engl. *Negative Likelihood Ratio* – LR^-)	Koliko puta je vjerojatnije da će model pogrešno klasificirati pozitivni slučaj kao negativan u odnosu na ispravnu klasifikaciju negativnog.	$\mathrm{LR}^- = \frac{FNR}{TNR}$ (42)
14	Dijagnostički omjer (*engl. Diagnostic Odds Ratio – DOR*)	Omjer koji izražava odnos između vjerojatnoće ispravne klasifikacije i pogrešne klasifikacije; veće vrijednosti ukazuju na bolju diskriminaciju. Za svaku klasu DOR se izračunava zasebno prema svim ostalim klasama.	$DOR_i = \frac{TP_i \cdot TN_i}{FP_i \cdot FN_i}$ (43)

Izvor: izradio autor prema (James et al., 2021).

Za svaku od sedam klasa zasebno su određeni TP, FP, FN i TN postupkom „jedna klasa prema svim ostalima". DOR je najprije izračunat za svaku klasu kao $DOR_i = (TP_i \cdot TN_i)/(FP_i \cdot FN_i)$, a sažeta vrijednost prikazana za model aritmetički je prosjek sedam klasnih DOR vrijednosti. Stoga se ona tumači kao opisni sažetak tih usporedbi, a ne kao zasebno definirani DOR za višeklasni ishod.

Značaj varijabli u modelima ključan je za interpretaciju rezultata. Ako neka varijabla ima visoku težinsku vrijednost u modelu, a model postiže visoke vrijednosti metrika, može se zaključiti da ta varijabla značajno doprinosi predikciji.
Kod modela potpornih vektora svaka ulazna varijabla pridonosi definiranju granica koje razdvajaju klase u višedimenzionalnom prostoru. Značajnost varijabli određuje se analizom težinskih vektora pridruženih pojedinačnim značajkama, pri čemu apsolutne vrijednosti tih koeficijenata pokazuju koliko svaka varijabla doprinosi razdvajanju klasa.

Varijable s većim apsolutnim vrijednostima koeficijenata imaju snažniji utjecaj na formiranje granica između klasa, što znači da model više računa na njih prilikom donošenja odluke o klasifikaciji. Ako varijabla s velikom težinskom vrijednošću doprinosi jasnom razdvajanju klasa, model će lakše ispravno klasificirati instance, posebno kod složenih ili rijetkih kategorija. Na taj način analiza težinskih vektora omogućuje identifikaciju ključnih prediktora koji poboljšavaju performanse modela i pruža uvid u koje varijable su najvažnije za precizno predviđanje rizika u obalnoj plovidbi.

Kod modela slučajnih šuma važnost varijabli određuje se kroz analizu doprinosa svake varijable smanjenju nečistoće čvorova u stablu. Svako stablo u šumi donosi odluke o klasifikaciji na temelju podjela čvorova koje koriste različite varijable. Varijable koje češće sudjeluju u ranijim i informacijski bogatim podjelama doprinose većem smanjenju nečistoće i stoga dobivaju veću važnost. Osim toga, koristi se permutacijska metoda smanjenja točnosti, pri kojoj se vrijednosti pojedine varijable nasumično miješaju, a zatim se mjeri pad performansi modela. Ako permutacija značajno smanjuje metrike performansi, poput točnosti, preciznosti ili F1-mjere, varijabla se smatra ključnom za predikciju. Ovakva analiza omogućuje identifikaciju najvažnijih varijabli koje model koristi za ispravnu klasifikaciju, uključujući i rijetke ili kritične klase rizika u obalnoj plovidbi. Varijable s većim značajem izravno doprinose poboljšanju metrika performansi jer omogućuju preciznije i pouzdanije razdvajanje klasa. Time se osigurava da model bolje razlikuje različite scenarije, što je posebno važno u kontekstu sigurnosnih procjena i donošenja odluka na temelju predikcija modela slučajnih šuma.
Garsonova metoda kod neuronskih mreža omogućava razlaganje težinskih vrijednosti veza kroz sve putanje od ulaznog do izlaznog sloja, čime se kvantificira doprinos svake ulazne varijable konačnoj odluci mreže. Svaka varijabla prolazi kroz mrežu slojeva, a težine veza određuju koliko signal iz ulaznog neurona utječe na aktivaciju skrivenih i izlaznih neurona te na konačnu klasifikaciju. Metoda zbraja apsolutne vrijednosti svih putanja od ulaznog do izlaznog neurona i normalizira ih u odnosu na ukupnu sumu težinskih putanja, čime se dobiva relativna težinska vrijednost varijable. Varijable s većom težinskom vrijednošću imaju snažniji utjecaj na aktivaciju mreže i konačnu predikciju, dok one s manjom doprinose manje. Relativna težinska vrijednost iz Garsonove metode direktno odražava koliko pojedina varijabla doprinosi točnosti, preciznosti, osjetljivosti i drugim ključnim metrikama performansi neuronske mreže. Varijable s visokim težinskim vrijednostima ključne su za pravilnu klasifikaciju, jer model uči da upravo te značajke najviše razlikuju različite klase ishoda. One često djeluju

u kombinaciji s drugim značajnim varijablama, što ilustrira njihovu međuzavisnost i doprinos ukupnoj prediktivnoj snazi modela. Analiza težinskih vrijednosti omogućava identifikaciju ulaznih varijabli koje su najvažnije za predikciju i daje uvid u strukturu odluka neuronske mreže (Garson, 1991; Beck, 2018).

Nakon provedene klasifikacije, modeli strojnog učenja korišteni su za procjenu vjerojatnosti pojave pojedinih kategorija glavnih događaja, čime se dobiva raspodjela vjerojatnosti po svim kategorijama ishoda. Budući da vjerojatnosti same po sebi ne predstavljaju rizik, u radu se rizik definira kao funkcija vjerojatnosti i težine posljedice, pri čemu se za svaki zapis izračunava raspodjela vjerojatnosti (čiji je zbroj jednak jedan) te kombinira s pripadajućim težinskim faktorima, čime se dobiva kvantitativna procjena rizika.

Model potpornih vektora temelji se na konstrukciji razdvajajućih granica u višedimenzionalnom prostoru značajki, pri čemu se, zbog izostanka izravne procjene vjerojatnosti, primjenjuje postupak kalibracije izlaza modela pomoću funkcija povezivanja (npr. sigmoidne ili *softmax* transformacije). Slučajne šume procjenjuju vjerojatnosti kao udio stabala koja klasificiraju zapis u određenu klasu, čime se dobiva robusna empirijska procjena temeljena na ansamblu. Kod neuronskih mreža vjerojatnosti proizlaze iz aktivacije izlaznog sloja, gdje se primjenom *softmax* funkcije dobiva normalizirana raspodjela vjerojatnosti nad svim klasama.

Radi povećanja robusnosti i prediktivne snage primijenjen je ansambl pristup temeljen na većinskom glasovanju između pojedinačnih modela, čime se smanjuje varijanca i poboljšava klasifikacija rjeđih kategorija. Dodatno, implementiran je slojeviti pristup u kojem se izlazi baznih modela kombiniraju pomoću meta-modela (slučajne šume), čime se integriraju komplementarne prednosti različitih klasifikatora i povećava točnost predikcije, osobito za slabije zastupljene klase. Rezultati su evaluirani matricom zabune i skupom standardnih metrika performansi, uključujući dijagnostički omjer vjerojatnosti (DOR), te pohranjeni radi transparentnosti i reproduktivnosti (James et al., 2021; Garson, 1991; Beck, 2018).

Ovakav integrirani pristup smanjuje rizik prenaučenosti, povećava stabilnost predikcija i omogućuje pouzdanu identifikaciju značajnih varijabli za procjenu rizika u obalnoj plovidbi, uz metodološku konzistentnost kroz primjenu istih ulaznih podataka i kriterija evaluacije.

3.6. Opis razvoja i treniranje modela

Sve varijable, uključujući numeričke i nominalne, tretirane su kao faktorske, čime je omogućena jednaka obrada podataka unutar svih modela. Prvo je pripremljen skup podataka tako što su sve nedostajuće vrijednosti unutar svake klase glavnog događaja popunjene najčešćim vrijednostima. Ovaj postupak osigurava konzistentnost i smanjuje mogućnost pogrešaka tijekom treniranja modela. Nakon toga, sve varijable pretvorene su u faktorske, što omogućava modelima da ispravno interpretiraju kategorijske vrijednosti.

3.6.1. Opis postupka treniranja modela potpornih vektora

Treniranje modela potpornih vektora započinje na pripremljenom trening skupu, gdje su sve varijable već pretvorene u faktorske, a nedostajući podaci popunjeni najčešćim vrijednostima.

Prvi korak u treniranju je definicija skupa vrijednosti hiperparametra *cost* (C):

$$cost_values < - c(61, 65, 67, 69, 73, 75, 77) \quad (44)$$

Navedene vrijednosti predstavljaju unaprijed zadani skup kandidata u rasponu od 61 do 77. Za svaku vrijednost model je treniran na trening skupu, a konačni odabir proveden je prema najvećoj točnosti na validacijskom skupu. Prema tom kriteriju odabrana je vrijednost C = 65.

Parametar *cost* kontrolira kompromis između minimiziranja pogrešaka na trening skupu i sposobnosti modela da generalizira na nove podatke. Visoka vrijednost parametra *cost* prisiljava model da što manje pogrešno klasificira trening primjere, čime se povećava točnost na poznatim podacima, ali može doći do prenaučenosti i smanjenja sposobnosti generalizacije na testni skup. S druge strane, niža vrijednost parametra *cost* dopušta više pogrešaka na trening skupu, što može smanjiti prenaučenost i poboljšati sposobnost modela da generalizira na nove, nepoznate primjere.

U ovom istraživanju provedena je ograničena pretraga sedam unaprijed zadanih vrijednosti parametra *cost*, u rasponu od 61 do 77. Ispitane vrijednosti omogućuju usporedbu točnosti klasifikacije pri različitim razinama penalizacije pogrešaka. U ovom slučaju odabrano je 7 konkretnih vrijednosti jer omogućuje dovoljan broj pokušaja za pronalazak optimalne postavke, a da pritom treniranje ostane računalno izvedivo; teoretski se moglo odabrati i više, npr. 93 ili više, no to bi značajno povećalo vrijeme treniranja, a ne bi nužno poboljšalo rezultate.

Za svaku od ovih sedam vrijednosti model se trenira i evaluira na validacijskom skupu. Ovo znači da se model ukupno trenira 7 puta, jednom za svaku vrijednost *cost*, čime se analizira kako promjena ovog hiperparametra utječe na performanse modela. Optimalna vrijednost *cost* odabire se na temelju najbolje kombinacije točnosti i stabilnosti modela na validacijskom skupu. Nakon što se odredi optimalna vrijednost, konačni model trenira se na spojenom trening i validacijskom skupu kako bi se iskoristio maksimalni broj primjera. Time se poboljšava učenje modela i osigurava da model u konačnoj verziji ima najbolje moguće performanse prije testiranja. Testiranje konačnog modela provodi se na odvojenom testnom skupu, koji nije korišten ni u treniranju ni u odabiru parametra, čime se osigurava nepristrana i objektivna procjena performansi modela.

Za dodatnu analizu važnosti varijabli koristi se linearni model potpornih vektora. Težine pridružene varijablama koriste se za određivanje njihovog relativnog doprinosa u klasifikaciji, odnosno za identifikaciju koje varijable najviše utječu na predviđanje klase glavnog događaja. Dobiveni rezultati sortirani su i spremljeni za daljnju analizu i vizualizaciju, što omogućava precizno uočavanje ključnih varijabli u modelu.

3.6.2. Opis treniranja modela slučajnih šuma i optimizacija broja stabala

Model slučajnih šuma temelji se na skupu velikog broja stabala odlučivanja koja zajednički donose konačnu odluku o klasi putem mehanizma glasanja. Svako pojedinačno stablo trenira se na nasumično odabranom podskupu podataka (*bootstrap uzorak*) te na nasumičnom podskupu značajki, čime se smanjuje međusobna korelacija stabala. Ovakav postupak značajno smanjuje varijancu modela u odnosu na pojedinačna stabla odlučivanja i povećava njegovu robusnost na šum i varijacije u podacima.

Ključni hiperparametar kod modela slučajnih šuma je *ntree*, odnosno broj stabala uključenih u skup. Broj stabala izravno utječe na stabilnost i pouzdanost predikcija: povećanjem vrijednosti *ntree* smanjuje se varijanca modela jer se konačna odluka temelji na prosjeku većeg broja stabala. Međutim, nakon određene točke povećanje broja stabala donosi sve manji dobitak u točnosti, dok se istovremeno povećavaju računalni zahtjevi i vrijeme treniranja modela.

U ovom istraživanju isprobane su vrijednosti parametra *ntree*: 120, 300, 360, 480 i 500 stabala. Ove vrijednosti odabrane su kako bi se obuhvatio raspon od relativno manjeg broja stabala, kod kojeg model još može biti osjetljiv na varijacije u podacima, do većeg broja stabala, gdje se očekuje stabilizacija performansi. Takav izbor omogućuje analizu kako se ponašanje modela mijenja s postupnim povećanjem kompleksnosti skupa. Iako je teorijski moguće koristiti i znatno veći broj stabala, praktična iskustva i literatura pokazuju da se nakon nekoliko stotina stabala dobitak u točnosti najčešće zanemarivo povećava.

Za svaku od navedenih vrijednosti parametra *ntree* model se trenira zasebno, što znači da se model trenira ukupno pet puta. Performanse svakog modela procjenjuju se na validacijskom skupu, a optimalna vrijednost *ntree* odabire se na temelju najbolje ravnoteže između točnosti i stabilnosti predikcija. Ovakav postupak optimizacije hiperparametara predstavlja općenito prihvaćen pristup u treniranju ansambl modela i ne odnosi se isključivo na ovaj konkretni skup podataka.

Nakon odabira optimalnog broja stabala, konačni model trenira se na spojenom trening i validacijskom skupu, čime se koristi maksimalna količina dostupnih podataka za učenje. Performanse tako treniranog modela zatim se ocjenjuju na potpuno odvojenom testnom skupu, koji nije korišten ni u fazi treniranja ni u procesu optimizacije hiperparametara, čime se osigurava nepristrana i objektivna procjena uspješnosti modela.

Važnost varijabli u modelu slučajnih šuma određuje se na temelju njihovog doprinosa u smanjenju nečistoće čvorova, pri čemu se najčešće koristi Gini indeks. Gini indeks mjeri stupanj nečistoće čvora u stablu odlučivanja, odnosno vjerojatnost pogrešne klasifikacije nasumično odabranog primjera koji pripada tom čvoru. Vrijednosti Gini indeksa kreću se od 0 do 1, pri čemu niža vrijednost označava veću homogenost čvora i jasnije razdvajanje klasa. Varijable koje najviše doprinose smanjenju Gini indeksa tijekom izgradnje stabala smatraju se najvažnijima za klasifikaciju, a njihova analiza omogućuje identifikaciju ključnih značajki koje imaju najveći utjecaj na predviđanje klase glavnog događaja.

3.6.3. Opis treniranja modela neuronske mreže i optimizacija hiperparametara

Treniranje modela neuronske mreže započinje definiranjem ključnih hiperparametara koji imaju presudan utjecaj na sposobnost mreže da uči obrasce iz podataka i generalizira naučeno znanje na nove primjere. Za razliku od jednostavnijih modela, neuronske mreže imaju velik broj slobodnih parametara, zbog čega je pravilna optimizacija hiperparametara nužna kako bi se izbjegla prenaučenost ili nedovoljno učenje modela.

U ovom radu optimiziraju se dva najvažnija hiperparametra neuronske mreže:

- *broj neurona u skrivenom sloju* (*size_values*), koji određuje kapacitet mreže za učenje nelinearnih odnosa između ulaznih varijabli i ciljne klase. Manji broj neurona ograničava sposobnost modela da uhvati složene obrasce, dok veći broj neurona omogućava modeliranje kompleksnijih odnosa, ali istovremeno povećava rizik od prenaučenosti i produžuje vrijeme treniranja.
- *parametar regularizacije* (*decay_values*), koji uvodi kaznu za velike vrijednosti težina u mreži. Time se sprječava da pojedine varijable previše dominiraju u procesu odlučivanja, što doprinosi stabilnijem učenju i boljoj sposobnosti generalizacije.

Vrijednosti hiperparametara nisu odabrane proizvoljno. Brojevi neurona u skrivenom sloju (6, 7 i 8) odabrani su na temelju preliminarnih testiranja i veličine skupa podataka. Budući da se radi o umjereno složenom problemu s ograničenim brojem primjera, korištenje znatno većeg broja neurona (npr. 20 ili 50) povećalo bi rizik od prenaučenosti bez značajnog dobitka u točnosti. Slično tome, vrijednosti parametra *decay* (0,5; 0,6; 0,7; 0,8) odabrane su kako bi se ispitao raspon umjerene do jače regularizacije, što je uobičajena praksa kod neuronskih mreža treniranih na realnim i heterogenim podacima.

Teoretski, raspon hiperparametara mogao je biti širi, a vrijednosti su mogle uključivati i veće brojeve (npr. 10 neurona ili *decay* = 0,9). Međutim, u praksi se hiperparametri biraju u razumnom rasponu koji odgovara složenosti problema i veličini skupa podataka, kako bi se izbjeglo nepotrebno povećanje računalnog opterećenja i nestabilnost modela. Stoga je ovaj izbor specifičan za promatrani problem, ali se temelji na općim principima treniranja neuronskih mreža.

Za pronalazak optimalne kombinacije hiperparametara koristi se ugniježđena petlja, kojom se sustavno testiraju sve moguće kombinacije vrijednosti *size_values* i *decay_values*. U ovom slučaju mreža se trenira ukupno 12 puta (3 vrijednosti broja neurona × 4 vrijednosti parametra regularizacije). Svaka kombinacija trenira se na trening skupu, a zatim se njezina uspješnost procjenjuje na validacijskom skupu. Ovaj postupak omogućava objektivnu usporedbu modela i odabir kombinacije koja daje najbolju ravnotežu između točnosti i stabilnosti.

Maksimalan broj iteracija treniranja postavljen je na 2000 kako bi se mreži omogućilo dovoljno vremena za konvergenciju, dok je ispis tijeka treniranja isključen radi preglednijeg izvođenja koda i smanjenja nepotrebnog ispisa.

Zbog velikog broja ulaznih varijabli i njihove moguće međusobne korelacije, neuronske mreže ponekad imaju poteškoća s konvergencijom. U takvim slučajevima primjenjuje se

postupak selekcije varijabli. Najprije se koristi model slučajnih šuma za rangiranje varijabli prema njihovoj važnosti. Varijable s najmanjim doprinosom postupno se uklanjaju, a treniranje neuronske mreže ponavlja se nakon svake redukcije skupa značajki. Ovaj iterativni postupak provodi se sve dok se ne postigne stabilna konvergencija modela. Ako problemi i dalje postoje, postupak se dodatno ponavlja uz precizniju eliminaciju najmanje informativnih varijabli. Važnost varijabli za ovu redukciju određena je na trenirajućem skupu, bez uporabe testnog skupa.
Kada se identificira funkcionalan i informativan podskup varijabli, neuronska mreža se ponovno trenira na reduciranom skupu podataka, koristeći optimalno odabrane hiperparametre i produženi broj iteracija. Završna evaluacija modela provodi se na zasebnom testnom skupu, koji nije korišten ni u treniranju ni u optimizaciji hiperparametara, čime se osigurava nepristrana i realna procjena performansi.

Ovim sustavnim postupkom treniranja, optimizacije hiperparametara i selekcije varijabli osigurava se da neuronska mreža maksimalno iskoristi dostupne podatke, uz istovremeno smanjenje rizika od prenaučenosti. Paralelna analiza važnosti varijabli dodatno doprinosi razumijevanju ponašanja modela i identificiranju ključnih značajki koje najviše utječu na predviđanje klase glavnog događaja.

3.6.4. Opis ansambl i slojevitog modela

Pored pojedinačnih modela (potpornih vektora, slučajnih šuma i neuronskih mreža), primijenjeni su ansambl ili skupni pristupi, s posebnim naglaskom na slojeviti model. Ansambl modeli kombiniraju predikcije više osnovnih modela kako bi se dobila konačna klasifikacija. Cilj ovog pristupa je povećanje stabilnosti i točnosti modela, posebno kada pojedinačni modeli pokazuju različite, komplementarne prednosti (Zhou, 2021).

Kod ansambl modela, predikcije svih osnovnih modela za svaku instancu testnog skupa se prikupljaju, a konačna klasa se određuje većinskim glasovanjem. Time se smanjuje varijanca predikcija jer pogreške pojedinačnih modela često ne koreliraju. Matrica zabune se potom formira na temelju ovih kombiniranih predikcija, što omogućuje procjenu učinkovitosti ansambl modela u odnosu na pojedinačne modele. Ovakva metoda je jednostavna, ali učinkovita, jer kombinira snagu različitih algoritama i često poboljšava metrike performansi poput točnosti, F1-mjere i osjetljivosti (Mishra et al., 2025).

Kod slojevitog pristupa, predikcije osnovnih modela na trening ili validacijski skup ne koriste se direktno za finalnu klasifikaciju, već služe kao ulazni podaci za meta-model. U ovom radu, meta-model je slučajna šuma koja uči kako optimalno kombinirati predikcije osnovnih modela za poboljšanje točnosti klasifikacije. Proces funkcionira u nekoliko koraka:

1. Trening osnovnih modela – Potporni vektori, slučajne šume i neuronske mreže treniraju se na trening skupu, pri čemu se određuju optimalni hiperparametri svakog modela (*cost* za potporne vektore, *ntree* za slučajne šume, *size* i *decay* za neuronsku mrežu).

2. Predikcije osnovnih modela – Svaki osnovni model generira predikcije na validacijskom skupu.
3. Kreiranje meta-skupa – Predikcije osnovnih modela koriste se kao nove varijable u meta-skupu. Svaka kolona predstavlja predikcije jednog osnovnog modela. Ciljna varijabla ostaje ista (stvarna klasa).
4. Trening meta-modela – Meta-model (ovdje slučajna šuma) uči kako kombinirati predikcije osnovnih modela kako bi se poboljšala klasifikacijska točnost. Model analizira koje kombinacije predikcija doprinose najboljoj odluci.
5. Primjena na testni skup – Nakon treniranja, slojeviti model se primjenjuje na predikcije osnovnih modela na testnom skupu. Konačne predikcije meta-modela zatim se koriste za kreiranje matrice zabune i izračun svih metrika performansi.

Prednost slojevitog modela je u tome što omogućuje povećanu sposobnost generalizacije na nepoznate podatke, smanjuje rizik od preučenja pojedinačnih modela i omogućuje identificiranje komplementarnih informacija koje različiti algoritmi prepoznaju (Zhou, 2021). Matrice zabune i pripadajuće metrike služe isključivo za evaluaciju performansi meta-modela i ne koriste se direktno pri kreiranju modela.

Kod ansambl pristupa procjena vjerojatnosti ne temelji se na pojedinačnom modelu, već na kombinaciji više modela s ciljem povećanja robusnosti i točnosti rezultata. U ovom istraživanju ansambl je formiran objedinjavanjem izlaza triju modela, metode potpornih vektora, slučajnih šuma i neuronskih mreža, pri čemu se konačna procjena vjerojatnosti dobiva njihovom agregacijom.

Za svaku instancu najprije se iz svakog pojedinog modela dobiva vektor vjerojatnosti nad sedam klasa. Konačna ansambl vjerojatnost zatim se određuje agregacijom tih vektora, najčešće:

- aritmetičkim prosjekom vjerojatnosti po klasama ili,
- ponderiranim prosjekom, pri čemu se modeli s boljim performansama vrednuju većim težinama.

Time se smanjuje utjecaj pojedinačne pogreške modela, a konačna raspodjela vjerojatnosti odražava konsenzus više algoritama. Ovakav pristup posebno je pogodan u sigurnosnim analizama jer ublažava preoptimiranost pojedinih modela i osigurava stabilniju procjenu rizika.

Slojeviti pristup predstavlja napredniji oblik ansambla u kojem se izlazi više baznih modela koriste kao ulazi dodatnog, meta-modela. U prvom stupnju treniraju se pojedinačni modeli: potpornih vektora, slučajne šume i neuronske mreže. Svaki od njih za svaki zapis generira vjerojatnosti za sedam glavnih događaja. Dobiveni vektori vjerojatnosti zatim se ne agregiraju izravno, već se spajaju u novu značajsku reprezentaciju.

U drugom stupnju trenira se meta-model (u ovom istraživanju model slučajnih šuma) koji uči kako kombinirati vjerojatnosti baznih modela kako bi se dobile što točnije konačne vjerojatnosti. Meta-model uči koje modele treba više slušati za pojedine klase, odnosno u kojim područjima podataka je koji algoritam pouzdaniji. Konačni rezultat slojevitog

pristupa je raspodjela vjerojatnosti na sedam kategorija glavnog događaja koja je optimizirana na temelju učenja iz pogrešaka baznih modela.
Primjenom ovih metodoloških postupaka dobiva se uvid ne samo u najvjerojatniju kategoriju glavnog događaja, nego i u razine rizika za sve moguće ishode nezgoda. Takva kvantitativna procjena vjerojatnosti ključna je za razumijevanje sigurnosnih aspekata obalne plovidbe, donošenje odluka i planiranje preventivnih mjera.
U programskom okruženju R, za implementaciju ansambl i slojevitog pristupa koriste se sljedeći paketi:

- *caret* – paket za sveobuhvatno strojno učenje, trening modela, validaciju i generiranje predikcija.
- *caretEnsemble* – paket specijaliziran za kreiranje i evaluaciju ansambl modela, uključujući slojeviti. Omogućuje jednostavno treniranje više osnovnih modela i meta-modela te izračun performansi.
- *openxlsx* – paket za učitavanje i spremanje podataka u Excel formatu, koristi se za dokumentiranje matrica zabune, važnosti varijabli i metrika performansi.

Ovim pristupom omogućeno je uspoređivanje individualnih modela s kombiniranim rješenjima, te identifikacija optimalne strategije za klasifikaciju nezgoda u obalnoj plovidbi. Ansambl i slojeviti modeli pružaju robustan okvir za integraciju komplementarnih informacija iz različitih algoritama, čime se postižu bolje performanse i stabilnije predikcije.

3.7. Višerazinski pristup validaciji modela

U svrhu provjere pouzdanosti i generalizacijske sposobnosti razvijenih modela provedena je dvostruka validacija rezultata.

U prvom koraku primijenjena je standardna validacija temeljena na podjeli podataka na trenirajući, validacijski i testni skup. Podaci su podijeljeni tako da je većina zapisa korištena za treniranje modela, dok su zasebni podskupovi korišteni za validaciju i završno testiranje. Na taj je način omogućena procjena točnosti modela na podacima koji nisu korišteni tijekom treniranja te smanjen rizik od prenaučenosti. Validacijski skup služio je za podešavanje ključnih hiperparametara modela, dok je testni skup korišten za konačnu procjenu performansi i robusnosti.

U drugom koraku provedena je analiza robusnosti modela s ciljem ispitivanja stabilnosti interpretacije te konzistentnosti procjene vjerojatnosti mogućih ishoda i posljedično procjene rizika.

Provedena je dodatna, eksperimentalna analiza na zasebnom validacijskom skupu podataka u kojem su ciljni događaji djelomično i kontrolirano modificirani. Time je simuliran scenarij prisutnosti pogrešno označenih podataka, nekonzistentnosti klasifikacije i promjena kriterija evidentiranja događaja, što predstavlja realističan problem u sigurnosnim i operativnim sustavima. Na taj je način ispitano:

- reagira li model u očekivanom smjeru kada se povećava ili smanjuje rizična varijabla,

- ostaju li raspodjele vjerojatnosti stabilne nakon promjene ciljnih oznaka,
- dolazi li do nekritične prilagodbe modela umjetno promijenjenim podacima.

Ova druga razina validacije predstavlja provjeru robusnosti i vjerodostojnosti modela, a ne samo njegove numeričke točnosti. Time se osigurava da model ne reproducira samo statističke obrasce prisutne u podacima, nego zadržava konzistentnost u skladu s pomorsko-nautičkom logikom i stvarnim sigurnosnim procesima u obalnoj plovidbi. Ovakav pristup omogućio je procjenu robusnosti modela, njegovu otpornosti na narušenu strukturu podataka te stabilnost procjene rizika u uvjetima djelomične nekonzistentnosti podataka.

Kombinacijom ove dvije validacije dobivena je pouzdana procjena prediktivne učinkovitosti modela te potvrđena stabilnost procjene vjerojatnosti pojave svakog od sedam glavnih događaja, što predstavlja osnovu za daljnju procjenu rizika. Za svaki primjer iz testnog skupa generirana je raspodjela vjerojatnosti po svim mogućim kategorijama glavnih događaja, koja je u kombinaciji s pripadajućim empirijskim indeksima posljedica omogućila kvantifikaciju ukupne vrijednosti rizika. Na taj način procjena rizika uzima u obzir istodobno vjerojatnost pojave pojedinog glavnog događaja i težinu njegovih posljedica.

3.8. Izvoz rezultata i dokumentiranje procesa

Ulazni podaci za analizu učitani su iz Excel datoteke u kojoj su prethodno pohranjeni svi relevantni podaci o promatranim varijablama. Svi rezultati modeliranja, uključujući numeričke prikaze poput matrica korelacije (10 najsnažnijih korelacija), metrike performansi i važnost varijabli, pohranjeni su u Excel datoteke. Grafički prikazi izvedeni su zasebno radi vizualne interpretacije rezultata, uključujući stupičaste dijagrame, usporedne grafove, i vizualizaciju neuronske mreže.

Za svaki od tri osnovna modela klasifikacije, potporni vektori, slučajne šume i neuronske mreže, generirane su predikcije na testnom skupu. Te predikcije predstavljaju klasifikaciju svake instance skupa podataka prema kategoriji glavnog događaja. Na temelju predikcija kreirane su matrice zabune, koje omogućuju usporedbu stvarnih i predviđenih vrijednosti te procjenu performansi modela. Svaka matrica zabune prikazuje broj pravilno klasificiranih instanci, broj pogrešno klasificiranih te raspodjelu pogrešaka po klasama, čime se dobiva detaljan uvid u prediktivnu moć svakog modela.
Izračunate su ključne metrike performansi, uključujući točnost, stopu greške, osjetljivost, specifičnost, preciznost, F1 mjeru, negativnu prediktivnu vrijednost, omjer dijagnostičkih vjerojatnosti i druge. Za svaki model izvedeni su i grafički prikazi važnosti varijabli: najsnažnijih 20 varijabli za potporne vektore i neuronsku mrežu, te sve varijable za slučajne šume. Dodatno, neuronska mreža vizualizirana je kroz sliku koja prikazuje sve slojeve, neurone i težine mreže, što omogućava detaljan uvid u strukturu i funkcionalnost mreže.

Za svaki od pet razvijenih modela (tri osnovna modela, ansambl model i slojeviti model) izračunate su vjerojatnosti pripadnosti zapisa za svih sedam definiranih glavnih

događaja. Vjerojatnosti su izvedene iz probabilističkih izlaza modela, pri čemu su kod potpornih vektora procjene dobivene transformacijom odluke modela u vjerojatnosti, kod slučajnih šuma kao relativni udio stabala koja predviđaju pojedinu klasu, dok su kod neuronskih mreža dobivene normalizacijom izlaznog sloja. Time je osigurana usporediva i konzistentna procjena vjerojatnosti pripadnosti klasama za sve modele.
Procjena izvedenog pokazatelja rizika provedena je odvojeno od procjene klasifikacijskih vjerojatnosti, u skladu s izrazima (15) – (20). Modelom procijenjene vjerojatnosti pripadnosti klasama (P_i) kombinirane su s empirijski određenim indeksom posljedica (C_i), izračunatim na temelju distribucije razina ozbiljnosti prema Tablici 4., čime je dobiven parcijalni doprinos riziku po varijablama i njihovim kategorijama. Agregacija rizika provedena je zbrajanjem parcijalnih doprinosa i izračunom ponderiranih vjerojatnosti, dok je konačna interpretacija provedena diskretizacijom u razine učestalosti i posljedica te njihovim pozicioniranjem unutar matrice rizika (Tablica 3.). Na taj način jasno je razdvojena probabilistička procjena od evaluacije rizika, uz očuvanje metodološke dosljednosti.

U ovom istraživanju pojam predviđanja odnosi se na predviđanje pripadnosti zapisa jednoj od sedam definiranih klasa glavnog pomorskog događaja na temelju njegovih poznatih obilježja. Modeli strojnog učenja treniraju se na povijesno evidentiranim događajima, dok se njihova prediktivna sposobnost procjenjuje na zapisima koji nisu korišteni za učenje modela. Na taj se način klasifikacija i predviđanje koriste kao povezani, ali metodološki različiti aspekti istog postupka: klasifikacija definira strukturu i pripadnost klasa, dok prediktivni model procjenjuje pripadnost novog, prethodno nepoznatog zapisa jednoj od tih klasa. Takav pristup ne predstavlja procjenu bezuvjetne vjerojatnosti nastanka nezgode u budućoj plovidbi, nego predviđanje vrste događaja uvjetovano time da je događaj evidentiran.

Sve predikcije, matrice zabune, vjerojatnosti pripadnosti klasama i izračunate metrike pohranjene su u Excel datoteke s odvojenim radnim listovima za svaki model, uključujući ansambl i slojeviti pristup, čime je omogućena transparentna analiza i usporedba rezultata.

Svi koraci modeliranja, validacije i evaluacije implementirani su u programskom okruženju R. Paketi *e1071*, *randomForest* i *nnet* korišteni su za treniranje osnovnih modela, uz podešavanje hiperparametara i optimizaciju modelskih struktura. Paket *caret* korišten je kao središnji okvir za particioniranje podataka, treniranje i validaciju modela, dok je *caretEnsemble* omogućio implementaciju ansambl i slojevitog pristupa kroz treniranje meta-modela nad izlazima baznih modela. Imputacija nedostajućih vrijednosti provedena je primjenom najčešće vrijednosti pomoću paketa *DMwR2*, dok su obrada, transformacija i agregacija podataka realizirane pomoću paketa *dplyr* i funkcija poput *group_by()*, *summarise()*, *mutate()* i *pivot_longer()*.

Učitavanje i pohrana podataka te rezultata modela provedeni su paketima *readxl* i *openxlsx*, dok su metričke performanse modela (točnost, preciznost, osjetljivost, F1-mjera i dr.) izračunate pomoću paketa *MLmetrics*. Vizualizacija rezultata, uključujući

prikaz važnosti varijabli i strukture neuronske mreže, realizirana je paketima *ggplot2* i *NeuralNetTools*, uz korištenje funkcija poput *ggplot()*, *geom_col()* i *geom_point()*.

Cjelokupan kod priložen je u dodatku rada, čime je osigurana potpuna transparentnost, ponovljivost i provjerljivost metodološkog postupka. Ovakav pristup omogućuje preciznu kvantifikaciju vjerojatnosti pripadnosti klasama, konzistentno izvođenje pokazatelja rizika te detaljan uvid u doprinos pojedinih varijabli ukupnoj vrijednosti izvedenog pokazatelja.

4. RAZVOJ MODELA I EVALUACIJA

U ovom poglavlju prikazani su rezultati provedene analize te validacija razvijenih modela strojnog učenja za procjenu rizika u obalnoj plovidbi. Prvi korak analize predstavlja korelacijska analiza varijabli, čiji je cilj bio ispitivanje međusobnih odnosa unutar pripremljenog skupa podataka u skladu s metodološkim okvirom prikazanim u Tablici 5.

4.1. Korelacijska analiza varijabli

Cilj ove faze bio je steći uvid u strukturu podataka, razumjeti međusobne povezanosti varijabli te osigurati kvalitetnu osnovu za interpretaciju rezultata i pouzdanu primjenu klasifikacijskih modela u nastavku istraživanja.

Na temelju Spearmanova koeficijenta korelacije identificirano je deset parova varijabli s najizraženijim međusobnim povezanostima, čiji su rezultati prikazani u Tablici 38.

Najizraženija pozitivna povezanost zabilježena je između varijabli *duljina* i *širina* (ρ = 0,87124). Ovakva visoka korelacija predstavlja očekivanu pojavu s obzirom na prirodu promatranih varijabli i njihove stvarne tehničke i fizikalne veze. Iako snažna povezanost može upućivati na preklapanje informacija, u ovom istraživanju se varijable ne isključuju, budući da njihov zajednički utjecaj može imati različitu interpretativnu i prediktivnu vrijednost u kasnijim fazama modeliranja. Naime, algoritmi strojnog učenja korišteni u analizi u stanju su različito vrednovati doprinos pojedinih varijabli ovisno o kontekstu i kombinaciji s drugim značajkama, pa se zadržavanjem svih relevantnih varijabli omogućuje potpunije učenje obrazaca i preciznija procjena rizika u složenim stvarnim uvjetima obalne plovidbe.

Značajna pozitivna povezanost uočena je i između varijable *bruto tonaža* i *duljina* (ρ = 0,60400), kao i između *bruto tonaža* i *širina* (ρ = 0,50399). Ovi odnosi odražavaju tehničke i konstrukcijske karakteristike brodova, budući da bruto tonaža raste s povećanjem dimenzija broda. Takva povezanost potvrđuje konzistentnost podataka i njihovu fizičku utemeljenost.

Visoka pozitivna korelacija zabilježena je i između varijabli *stanje mora* i *jačina vjetra* (ρ = 0,60086). Ovaj rezultat ukazuje na snažnu povezanost meteoroloških uvjeta, gdje jači vjetrovi najčešće dovode do pogoršanja stanja mora. Ovakav odnos ima poseban značaj u kontekstu procjene rizika u obalnoj plovidbi, jer kombinacija nepovoljnih vremenskih uvjeta izravno utječe na sigurnost plovidbe.

Pozitivna povezanost uočena je između varijabli *ozbiljnost nezgode* i *šteta na brodu* (ρ = 0,34003). Ovaj rezultat upućuje na to da karakter događaja može biti povezan s razinom nastale materijalne štete, što je logično s obzirom na to da određene vrste nezgoda inherentno nose veći rizik od oštećenja.

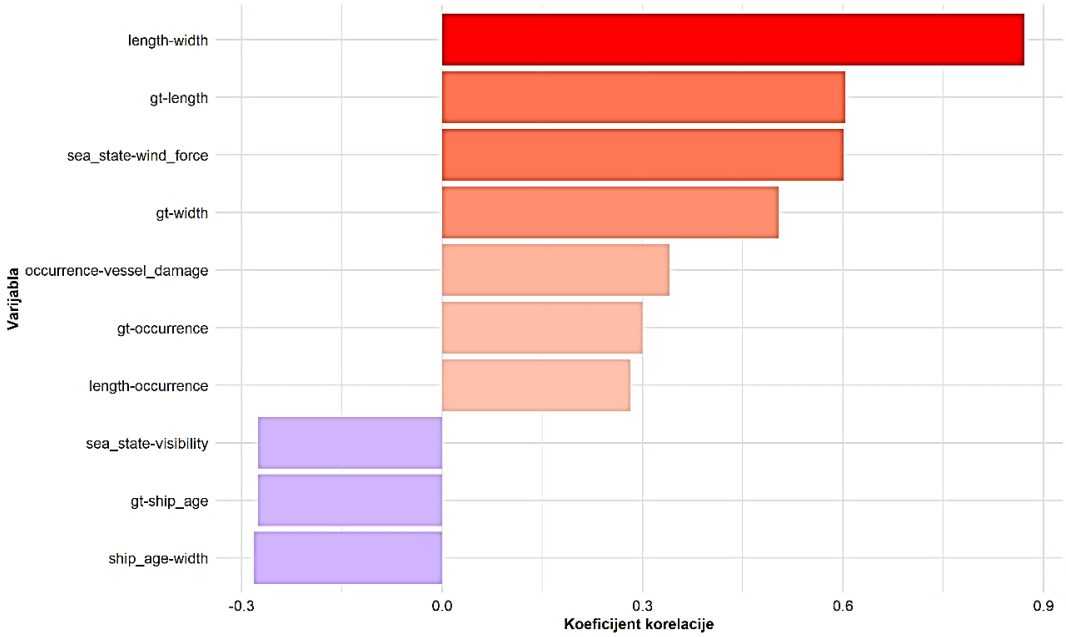
Bruto tonaža pokazuje pozitivnu povezanost s varijablom *ozbiljnost nezgode* (ρ = 0,30013), kao i *duljina* broda (ρ = 0,28203). Ovi odnosi mogu sugerirati da tehničke karakteristike broda, poput veličine i dimenzija, utječu na vjerojatnost pojave određenih tipova nezgoda ili na način na koji se nezgode manifestiraju.

Među deset najsnažnijih povezanosti uočene su i tri negativne korelacije. Negativna povezanost između varijabli *starost_broda* i *širina* (ρ = –0,28117), kao i između *bruto tonaža* i *starost broda* (ρ = –0,27560), može upućivati na to da su noviji brodovi u prosjeku većih dimenzija i veće tonaže. Također, može značiti da se stariji, manji i manje učinkoviti brodovi ranije povlače iz uporabe, dok veći brodovi ostaju dulje aktivni. Time su u uzorku ostali stariji, ali i dalje relativno veliki brodovi, što dodatno pojačava uočeni obrazac. Ovaj trend može odražavati tehnološki napredak, promjene u brodogradnji ili prilagodbu tržišnim zahtjevima. Negativna korelacija između *stanje mora* i *vidljivost* (ρ = –0,27537) potvrđuje očekivani utjecaj nepovoljnih vremenskih uvjeta na smanjenje vidljivosti, što predstavlja dodatni čimbenik rizika u plovidbi.
Vizualni prikaz deset najsnažnijih Spearmanovih korelacija među ordinalnim varijablama prikazan je na Slici 8., gdje su jasno istaknuti smjer i intenzitet povezanosti.

Tablica 38. Deset najsnažnijih Spearmanovih korelacija između ordinalnih varijabli

Par	Koeficijent
length – width	0,87124
gt – length	0,60400
sea_state – wind_force	0,60086
gt – width	0,50399
occurrence – vessel_damage	0,34003
gt – occurrence	0,30013
length – occurrence	0,28203
ship_age – width	-0,28117
gt – ship_age	-0,27560
sea_state – visibility	-0,27537

Izvor: izradio autor



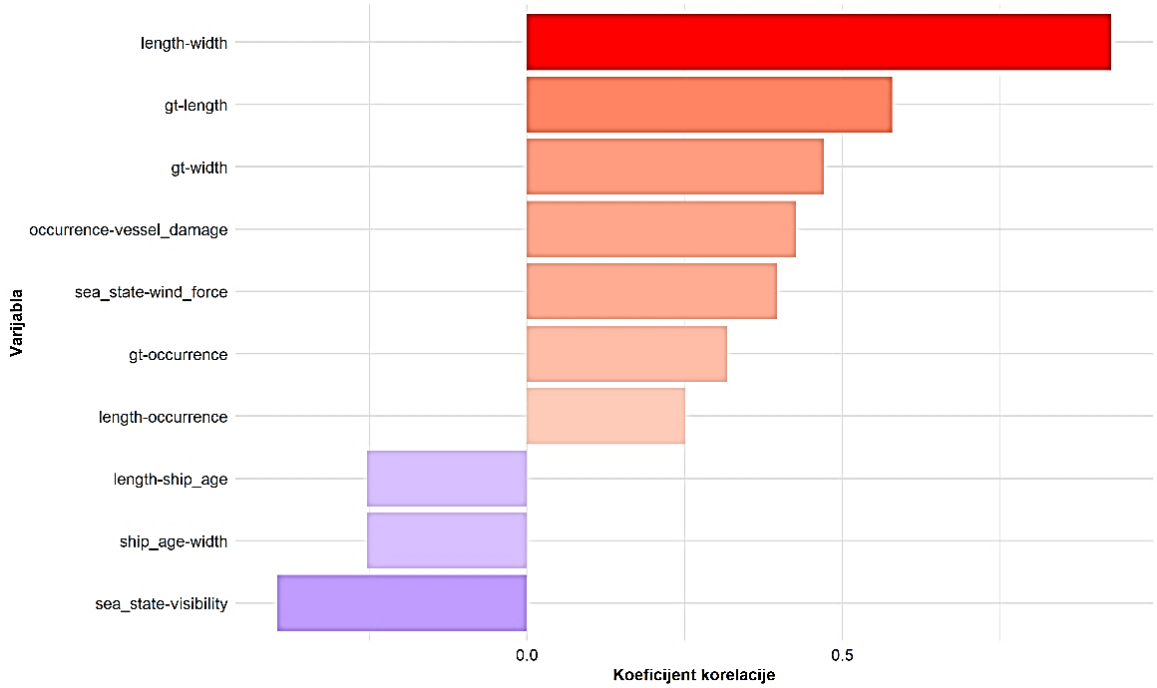
Slika 8. Deset najsnažnijih Spearmanovih korelacija među ordinalnim varijablama
Izvor: izradio autor

Visoke vrijednosti Pearsonovih koeficijenata korelacije prikazane u Tablici 39. potvrđuju očekivane odnose među promatranim varijablama, budući da se one odnose na međusobno povezane tehničke karakteristike brodova i okolišne čimbenike plovidbe. Posebno izražena povezanost između *duljine* i *širine* broda jasno je vidljiva i na Slici 9., čime se dodatno potvrđuje njihova proporcionalnost. Korelacije koje uključuju *bruto tonažu* odražavaju njezinu zavisnost o osnovnim dimenzijama broda, dok negativne povezanosti sa *starosti broda* mogu upućivati na strukturne promjene flote.

Tablica 39. Deset najsnažnih Pearsonovih korelacija između ordinalnih varijabli

Par	Koeficijent
length – width	0,92617
gt – length	0,57891
gt – width	0,47052
occurrence – vessel_damage	0,42629
sea_state – wind_force	0,39644
sea_state – visibility	-0,39589
gt – occurrence	0,31724
ship_age – width	-0,25368
length – ship_age	-0,25359
length – occurrence	0,25127

Izvor: izradio autor

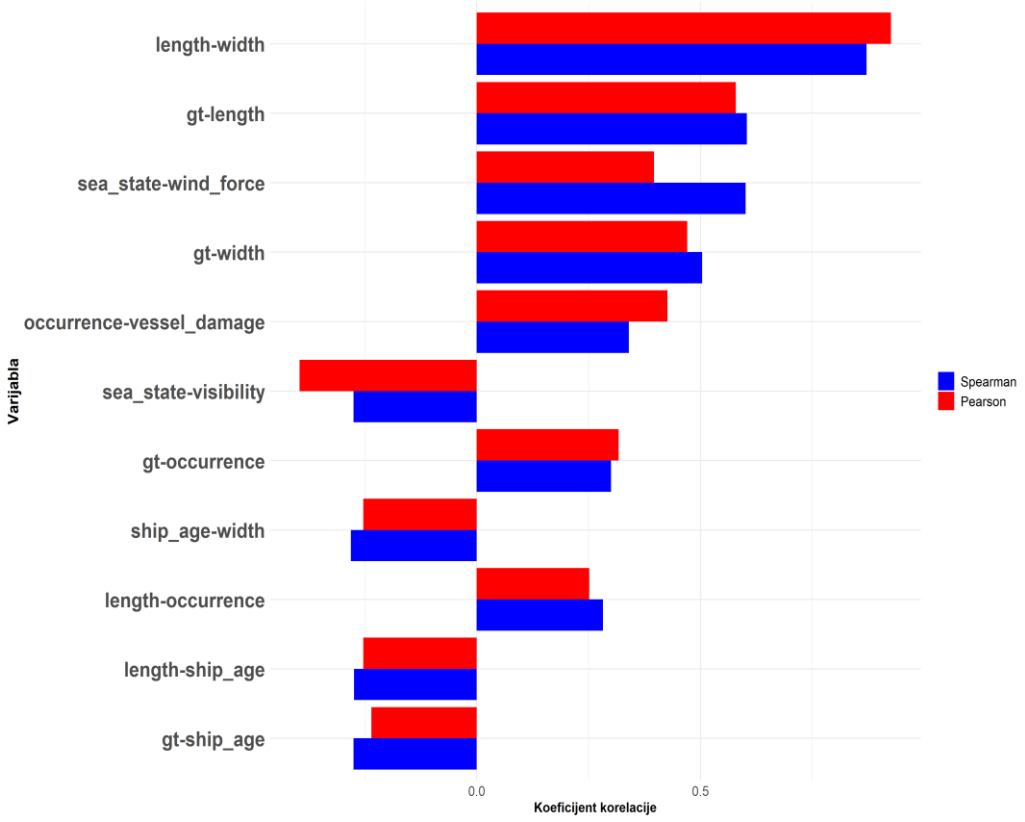


Slika 9. Deset najsnažnih Pearsonovih korelacija između ordinalnih varijabli

Izvor: izradio autor

Na temelju usporednog prikaza najsnažnijih korelacija prema Spearmanovu i Pearsonovu koeficijentu (Slika 10.) uočavaju se jasni i konzistentni obrasci koji potvrđuju valjanost korelacijske analize. Oba koeficijenta pokazuju vrlo visoku pozitivnu

povezanost između varijabli *duljina* i *širina* (ρ = 0,87124; r = 0,92617), što je očekivano zbog njihove proporcionalne povezanosti.
Snažna pozitivna korelacija prisutna je i između *bruto tonaže* i *duljine* (ρ = 0,60400; r = 0,57891) te *bruto tonaže* i *širine* (ρ = 0,50399; r = 0,47052), uz stabilan i približno linearan odnos. Umjerene korelacije, poput odnosa *ozbiljnost nezgode – oštećenja*, pokazuju nešto višu linearnu komponentu (r = 0,42629 vs. ρ = 0,34003), dok kod operativnih varijabli, primjerice *stanje mora – jačina vjetra*, Spearmanov koeficijent (ρ = 0,60086) nadmašuje Pearsonov (r = 0,39644), ukazujući na monotoni, ali nelinearan odnos.
Umjerena, ali konzistentne pozitivne korelacije prisutne su i između *bruto tonaže* i *ozbiljnost nezgode* (ρ = 0,30013; r = 0,31724) te *duljine* i *ozbiljnost nezgode* (ρ = 0,28203; r = 0,25127), što sugerira kumulativni utjecaj konstrukcijskih karakteristika na učestalost događaja.
Zaključno, Spearmanova i Pearsonova analiza pokazuju visoku usklađenost kod tehnički uvjetovanih odnosa, dok Spearman pruža dodatnu interpretacijsku vrijednost kod nelinearnih i okolišnih čimbenika, čime se omogućuje pouzdano modeliranje rizika u obalnoj plovidbi.



Slika 10. Usporedba Spearmanovih i Pearsonovih korelacija za deset najsnažnijih odnosa između ordinalnih varijabli

Izvor: izradio autor

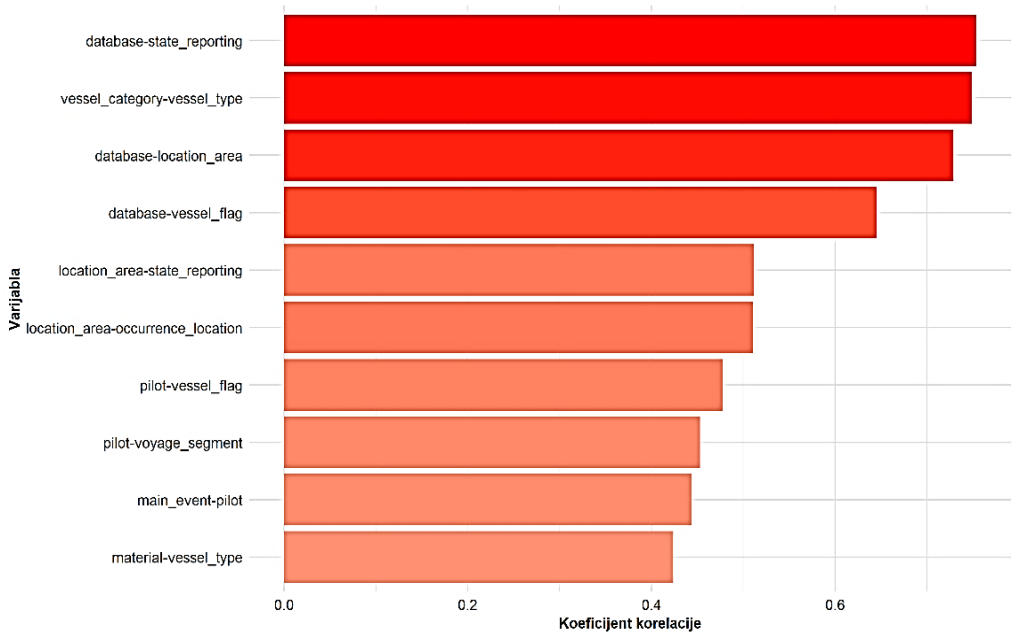
Tablica 40. i Slika 11. prikazuju deset najsnažnijih povezanosti između nominalnih varijabli prema Cramérovu V koeficijentu. Najizraženija povezanost uočava

se između *baza podataka* i *država izvjestiteljica* (V = 0,75413) te *kategorija broda* i *tip broda* (V = 0,74896), što odražava dosljednost u klasifikaciji brodova i sustavu izvještavanja. Posebno je značajna veza između *glavni događaj* i *peljar* (V = 0,44381), pokazuje da se raspodjela vrsta evidentiranih događaja razlikuje ovisno o prisutnosti peljara. Ovaj nalaz opisuje povezanost unutar analiziranog skupa nezgoda, dok sam koeficijent ne određuje povoljan ili nepovoljan doprinos peljara ishodu nezgode.

Tablica 40. Deset najsnažnijih Cramérovih V koeficijenata između nominalnih varijabli

Par	Koeficijent
database – state_reporting	0,75413
vessel_category – vessel_type	0,74896
database – location_area	0,72924
database – vessel_flag	0,64512
location_area – state_reporting	0,51128
location_area – occurrence_location	0,51063
pilot – vessel_flag	0,47786
pilot – voyage_segment	0,45319
main_event – pilot	0,44381
material – vessel_type	0,42387

Izvor: izradio autor



Slika 11. Deset najsnažnijih korelacija između nominalnih varijabli prema Cramérovu V koeficijentu

Izvor: izradio autor

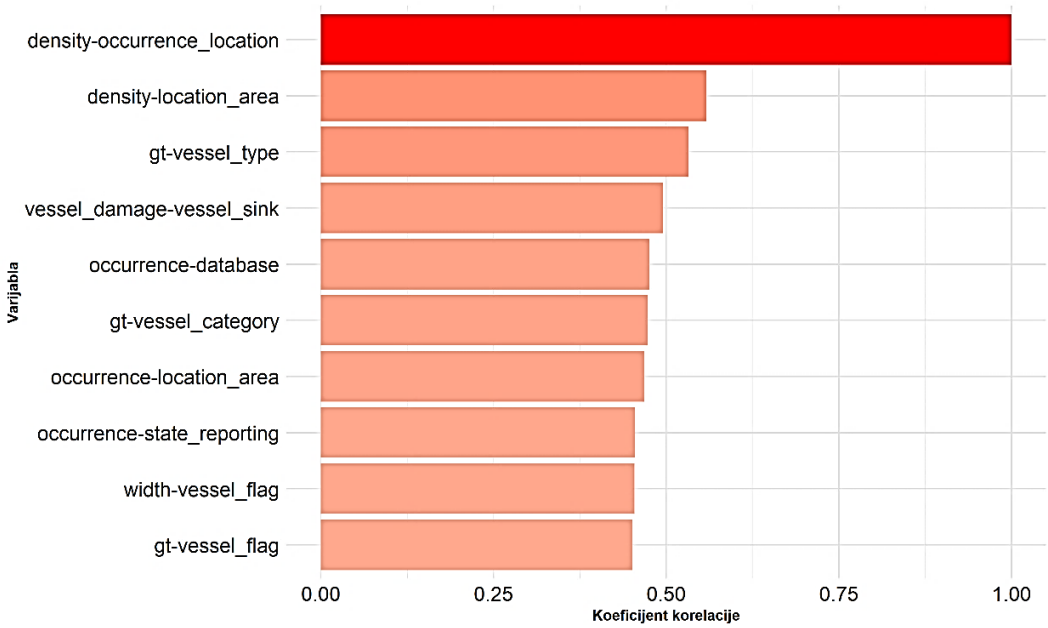
Tablica 41. i Slika 12. prikazuju deset najsnažnijih povezanosti između ordinalnih i nominalnih varijabli prema Cramérovu V koeficijentu. Najviša povezanost očituje se između *gustoća* i *mjesto pomorske nezgode* (V = 1,00000), što ukazuje na potpuno podudaranje vrijednosti i snažnu ovisnost prostornih i gustoćom definiranih varijabli.

Znatne korelacije prisutne su i između *gustoća* i *područje nezgode* (V = 0,55851), *bruto tonaža* i *tip broda* (V = 0,53212) te *oštećenje broda* i *potonuće broda* (V = 0,49561), što odražava logičnu povezanost tehničkih karakteristika i ishoda nezgode. Ostali parovi pokazuju umjerenu povezanost, primjerice *bruto tonaža* i *kategorija broda* (V = 0,47332) ili *širina* i *zastava broda* (V = 0,45406), potvrđujući da konstrukcijske i operativne karakteristike imaju statistički značajan utjecaj na nominalne varijable u sustavu evidencije. Ukupno, analiza ističe ključne strukturalne i operativne veze koje su relevantne za razumijevanje rizika u obalnoj plovidbi.

Tablica 41. Deset najsnažnijih Cramérovih V koeficijenata između ordinalnih i nominalnih varijabli

Par	Koeficijent
density – occurrence_location	1,00000
density – location_area	0,55851
gt – vessel_type	0,53212
vessel_damage – vessel_sink	0,49561
occurrence – database	0,47575
gt – vessel_category	0,47332
occurrence – location_area	0,46850
occurrence – state_reporting	0,45528
width – vessel_flag	0,45406
gt – vessel_flag	0,45104

Izvor: izradio autor



Slika 12. Deset najsnažnijih korelacija između ordinalnih i nominalnih varijabli prema Cramérovu V koeficijentu

Izvor: izradio autor

Korelacijska analiza ordinalnih i nominalnih varijabli u obalnoj plovidbi pokazala je jasne i očekivane obrasce povezanosti. Najjače korelacije među ordinalnim varijablama zabilježene su između *duljina* i *širina* te između *bruto tonaža* i dimenzija broda, dok su okolišne varijable poput *stanje mora* i *jačina vjetra* povezane s operativnim rizikom. Analiza nominalnih varijabli prema Cramérovu V istaknula je snažne veze u klasifikaciji brodova, s posebnom ulogom *glavni događaj – peljar*. Kombinirani odnosi ordinalnih i nominalnih varijabli, primjerice *gustoća – područje nezgode*, potvrđuju važnost tehničkih i operativnih karakteristika za procjenu rizika. Rezultati osiguravaju konzistentnu i pouzdanu osnovu za daljnje modeliranje i predikciju pomorskih nezgoda i procjenu rizika u obalnoj plovidbi.

Agregirani koeficijenti (C_i) prikazani u Tablici 43. izravno su dobiveni iz empirijske distribucije prikazane u Tablici 42., gdje su za svaki glavni događaj izračunati udjeli pojedinih razina ozbiljnosti te njihovi ponderirani doprinosi primjenom IMO FSA težina. Indeks C_i izračunat je iz empirijske raspodjele razina varijable *ozbiljnost nezgode* za pojedinu klasu *glavnog događaja*, uz primjenu odabranih težina posljedica. Stoga nije neovisna vanjska mjera posljedica: ista varijabla *ozbiljnost nezgode* korištena je i kao jedan od ulaza klasifikacijskog modela. Zbrajanjem tih doprinosa dobiven je jedinstveni indeks posljedice za svaki glavni događaj, čime je osigurana kvantitativna i usporediva mjera težine posljedica. Najviše vrijednosti (C_i) zabilježene su za događaje 1 i 6, što potvrđuje njihovu dominantnu ulogu u strukturi rizika, dok niže vrijednosti za događaje 5 i 7 ukazuju na relativno manji utjecaj posljedica.

Ovakav empirijski definiran indeks (C_i) predstavlja fiksnu komponentu u daljnjoj procjeni rizika, dok se druga komponenta, vjerojatnost pojave (P_i), procjenjuje primjenom modela strojnog učenja. Za svaki od pet razvijenih modela izračunate su vjerojatnosti pripadnosti svakog zapisa jednoj od sedam klasa glavnih događaja. Te vjerojatnosti dobivene su iz izlaza modela i potom agregirane po nezavisnim varijablama i njihovim kategorijama korištenjem funkcija *group_by()* i *summarise()*, čime su dobivene prosječne i reprezentativne vrijednosti (P_i) koje opisuju učestalost pojave događaja u ovisnosti o pojedinim čimbenicima.

Nakon toga provedeno je povezivanje vjerojatnosti i posljedica, pri čemu su podaci o (P_i) i (C_i) objedinjeni primjenom funkcije *left_join()*, a parcijalni doprinos riziku izračunat je za svaku kombinaciju varijable, kategorije i glavnog događaja pomoću funkcije *mutate()*, u skladu s izrazom (15). Time je dobiven (R_i), koji predstavlja kvantificirani doprinos pojedinog čimbenika ukupnom riziku. Ukupni rizik zatim je agregiran po varijablama i glavnim događajima zbrajanjem parcijalnih doprinosa, u skladu s izrazom (16), dok je dodatno izračunata i agregirana vjerojatnost kao ponderirani prosjek prema izrazu (20), čime su uvaženi relativni doprinosi pojedinih varijabli.

Dobivene kontinuirane vrijednosti vjerojatnosti i posljedica potom su transformirane u diskretne razine primjenom kriterija iz Tablice 4., gdje su (P) vrijednosti prevedene u razine učestalosti (F1 – F5), a (C_i) u razine posljedica (C1 – C4). Njihovom kombinacijom unutar matrice rizika (Tablica 3.) dobiven je konačni broj rizika kao umnožak razina

učestalosti i posljedica, čime je omogućena jasna klasifikacija rizika na mali, srednji i veliki.

Na razini implementacije, svi navedeni koraci realizirani su u programskom okruženju R. Vjerojatnosti su izračunate za svaki model i pohranjene u strukturirane Excel datoteke, nakon čega su transformirane u *long* format pomoću funkcije *pivot_longer()* radi analize po glavnim događajima. Izračuni (R_i), agregacije po varijablama te klasifikacija rizika provedeni su konzistentno za svih pet modela, čime je omogućena njihova izravna usporedba.

Na taj način uspostavljena je potpuna metodološka veza između empirijske distribucije posljedica (Tablica 42.), agregiranih indeksa (C_i) (Tablica 43.), modelom procijenjenih vjerojatnosti i konačne procjene rizika. Ovaj integrirani pristup omogućuje kvantitativno utemeljenu, transparentnu i konzistentnu analizu rizika, pri čemu se jasno identificiraju ključni čimbenici i njihov doprinos nastanku pomorskih nezgoda.

Tablica 42. Empirijska distribucija pojavnosti i ponderiranih doprinosa po glavnim događajima

Glavni događaj	Ozbiljnost nezgode	n	Udio	Težine	Doprinos
1	1	43	0,53750	10	5,37500
1	2	2	0,02500	1	0,02500
1	3	1	0,01250	0,1	0,00125
1	4	34	0,42500	0,01	0,00425
2	1	42	0,12000	10	1,20000
2	2	38	0,10857	1	0,10857
2	3	113	0,32286	0,1	0,03229
2	4	157	0,44857	0,01	0,00449
3	1	5	0,01805	10	0,18051
3	2	50	0,18051	1	0,18051
3	3	11	0,03971	0,1	0,00397
3	4	211	0,76173	0,01	0,00762
4	1	27	0,08385	10	0,83851
4	2	43	0,13354	1	0,13354
4	3	162	0,50311	0,1	0,05031
4	4	90	0,27950	0,01	0,00280
5	1	1	0,00270	10	0,02703
5	2	12	0,03243	1	0,03243
5	4	357	0,96486	0,01	0,00965
6	1	44	0,48889	10	4,88889
6	2	4	0,04444	1	0,04444
6	4	42	0,46667	0,01	0,00467
7	1	1	0,00980	10	0,09804
7	2	14	0,13725	1	0,13725
7	3	8	0,07843	0,1	0,00784
7	4	79	0,77451	0,01	0,00775

Izvor: izradio autor

Napomena: 1 – Prevrtanje/naginjanje; 2 – Sudar; 3 – Udaranje u nepomični objekt; 4 – Nasukavanje; 5 – Oštećenje/gubitak opreme; 6 – Prodor mora/potonuće; 7 – Gubitak upravljivosti.

Tablica 43. Agregirani koeficijent C_i po glavnim događajima

Glavni događaj	C_i
1	5,40550
2	1,34534
3	0,37260
4	1,02516
5	0,06911
6	4,93800
7	0,25088

Izvor: izradio autor

Sljedeća poglavlja prikazuju rezultate modela s naglaskom na klasifikaciju, vjerojatnosti i procjenu rizika, pri čemu se identificirane relacije između varijabli koriste za preciznu identifikaciju čimbenika rizika i razlikovanje stvarnih nezgoda u obalnoj plovidbi.

4.2. Rezultati klasifikacije, predikcije nezgoda i procjene rizika primjenom potpornih vektora

Model potpornih vektora implementiran je za klasifikaciju događaja koji predstavljaju sigurnosne rizike u obalnoj plovidbi. S obzirom na sedam klasa i značajnu međuovisnost varijabli, korištena je radijalna bazna funkcija koja omogućuje nelinearno razdvajanje klasa u višedimenzionalnom prostoru, čime se poboljšava prepoznavanje složenih i nelinearnih obrazaca ponašanja brodova i uvjeta plovidbe. Hiperparametar *cost (C)* koji kontrolira kompromis između penalizacije pogrešaka na trening uzorcima i širine margine razdvajanja, ispitan je prema postupku opisanom u poglavlju 3.6.1., testiran je s vrijednostima 61, 65, 67, 69, 73, 75 i 77. Za konačni model odabrana je vrijednost 65 zbog najbolje učinkovitosti na validacijskom skupu.

Analiza matrice zabune (Tablica 44.) pokazala je da model potpornih vektora najstabilnije klasificira klase *3* (*Udaranje u nepomični objekt*), *4* (*Nasukavanje*) i *5* (*Oštećenje / Gubitak opreme*). Konkretno, klasa *3* imala je 34 točno predviđena uzorka, uz nekoliko pogrešaka prema klasama *1*, *2*, *4*, *5* i *7*, što ukazuje na preklapanje obrazaca među sličnim vrstama događaja. Klasa *4*, 33 ispravno klasificirana uzorka, uz pogreške prema klasama *1*, *2*, *3*, *5* i *6*, što naglašava kompleksnost u razlikovanju njenih karakteristika. U klasi 5 zabilježen je najveći apsolutni broj točnih klasifikacija, njih 47, uz manji broj pogrešaka prema klasama *3* i *7*, što sugerira dijeljenje nekih značajki s tim kategorijama.

Klasa 2 (*Sudar*) pogrešno je klasificirana 13 puta, dok je u klasi 6 *(Prodor mora/Potonuće*) zabilježeno devet točnih predikcija, što odražava složenost tih događaja i djelomično preklapanje s drugim klasama. Ukupno, matrica zabune pokazala je

sposobnost modela da prepozna većinu događaja s dobrim odstupanjem u složenim i višeklasnim uvjetima, ali istovremeno ističe izazove u razlikovanju sličnih incidenata.

Tablica 44. Matrica zabune za model potpornih vektora

Stvarna klasa → / Predikcija ↓	1	2	3	4	5	6	7
1	8	1	1	2	0	1	0
2	0	32	2	10	0	0	1
3	2	5	34	2	3	1	2
4	1	11	2	33	1	2	1
5	0	0	2	0	47	0	7
6	1	2	0	1	0	9	0
7	0	1	0	0	4	0	4

Izvor: izradio autor

Ključne klasifikacijske metrike prikazane u Tablici 45. daju pregled uspješnosti modela potpornih vektora u razlikovanju sedam klasa glavnih događaja. Ukupna *točnost* (*Accuracy*) iznosi 0,70763, odnosno približno 70,8 % zapisa ispravno je klasificirano, dok *stopa pogreške* (*ErrorRate*) iznosi 0,29237, što upućuje na postojanje prostora za dodatna poboljšanja u razlikovanju klasa glavnih događaja.

Makroprosječna *osjetljivost* (*Sensitivity*) iznosi 0,66335 i opisuje sposobnost prepoznavanja zapisa koji pripadaju pojedinoj klasi. Makroprosječna *specifičnost* (*Specificity*) od 0,94971 opisuje sposobnost prepoznavanja zapisa koji ne pripadaju promatranoj klasi. Makroprosječna *preciznost* (*Precision*) iznosi 0,65891, dok *F1-mjera* iznosi 0,65636 i objedinjuje preciznost i osjetljivost.

Vrijednosti *FDR* = 0,34109 i *FOR* = 0,05085 predstavljaju makroprosječne udjele pogrešnih klasifikacija među pozitivnim odnosno negativnim predikcijama u pristupu pojedina klasa nasuprot ostalima. *Pozitivni omjer vjerodostojnosti* (*LR_Positive*) iznosi 19,18106, a *negativni omjer vjerodostojnosti* (*LR_Negative*) 0,35454. Ti pokazatelji dodatno opisuju razlikovanje pripadnosti pojedinoj klasi od nepripadnosti toj klasi. Makroprosječni *dijagnostički omjer izgleda* (*DOR*) iznosi 62,69772 te sažima odnos točnih i pogrešnih klasifikacija pojedinih klasa u odnosu na ostale klase, te potvrđuje snažnu ukupnu dijagnostičku moć modela potpornih vektora u razlikovanju klasa glavnih događaja u obalnoj plovidbi.

Tablica 45. Ključne klasifikacijske metrike modela potpornih vektora

Model	SVM
Accuracy	0,70763
ErrorRate	0,29237
Sensitivity (makroprosjek)	0,66335
Specificity	0,94971
Precision (makroprosjek)	0,65891
NPV	0,94915
F1 (makroprosjek)	0,65636
FDR	0,34109

FOR	0,05085
FPR	0,05029
FNR	0,33665
LR_Positive	19,18106
LR_Negative	0,35454
DOR (makroprosjek)	62,69772

Izvor: izradio autor

Ukupno, rezultati su pokazali da kombinacija tehničkih karakteristika broda, okolišnih čimbenika, operativnih te vremensko analitički kontekst pomorskih nezgoda omogućuje pouzdanu klasifikaciju potencijalnih *glavnih događaja*. Model se stoga može koristiti kao alat za podršku odlučivanju u obalnoj plovidbi, gdje su pravovremeno prepoznavanje potencijalno opasnih događaja i procjena njihove razine rizika ključni za unapređenje sigurnosti i optimizaciju reakcije na nezgode.

Analiza doprinosa varijabli u modelu potpornih vektora (Tablica 46. i Slika 13.) je pokazala da najznačajniji utjecaj na klasifikaciju nezgoda u obalnoj plovidbi ima varijabla *onečišćenjeNe*, što naglašava važnost prisutnosti i koncentracije onečišćenja u području plovidbe. Sljedeće po važnosti su prostorne i okolišne varijable, prvenstveno *gustoća prometa4* i *vidljivost4*, koje dodatno ističu utjecaj koncentracije događaja i uvjeta vidljivosti. Administrativne varijable (*baza podataka MAIB*, *država izvjestiteljica UK*, *područje nezgode UK*) također značajno doprinose modelu, što odražava kvalitetu i dosljednost sustava evidencije.

Tehnički čimbenici broda, poput *materijal trupa Čelik*, *kategorija broda Trgovački >=100GT* i *starost broda 5*, te operativne varijable (*ozbiljnost nezgode 3*, *mjesto pomorske nezgode Lučko područje*), dodatno oblikuju predikciju modela, potvrđujući povezanost konstrukcijskih karakteristika i lokacijskih uvjeta s vrstom *glavnog događaja*. U manjoj mjeri, okolišni čimbenici (*stanje mora 2*, *jačina vjetra 3*) i prisustvo *peljar Ne* pokazuju važnost stručne intervencije i meteoroloških uvjeta u sigurnosti plovidbe.

Sveukupno, model uspješno kombinira tehničke karakteristike broda te operativne, prostorne i administrativne čimbenike, omogućujući preciznu klasifikaciju glavnih događaja u složenim uvjetima obalne plovidbe. Procijenjene vjerojatnosti pojedinih glavnih događaja mogu se zatim, u kombinaciji s empirijski određenim indeksima posljedica, koristiti za daljnju kvantifikaciju razine rizika.

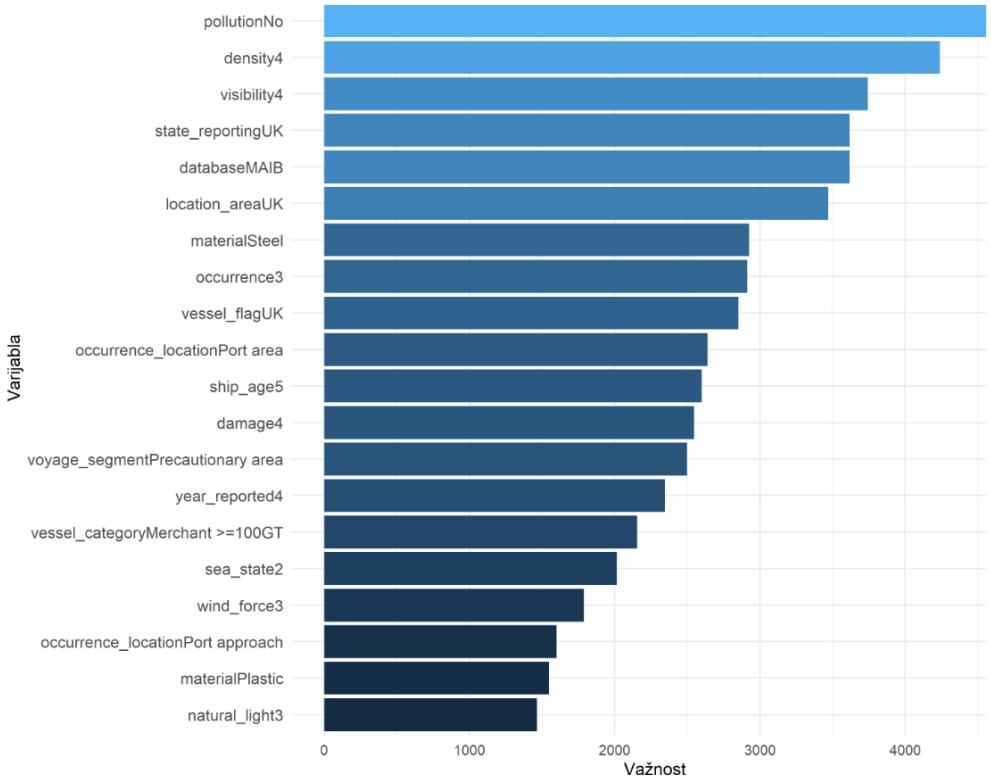
Tablica 46. Varijable najsnažnijeg utjecaja u modelu potpornih vektora

Varijabla	Važnost
pollutionNo	4557,332
density4	4238,153
visibility4	3743,064
databaseMAIB	3615,905
state_reportingUK	3615,905
location_areaUK	3470,106
materialSteel	2923,681

occurrence3	2912,191
vessel_flagUK	2850,659
occurrence_locationPort area	2638,204
ship_age5	2599,666
damage4	2546,937
voyage_segmentPrecautionary area	2496,740
year_reported4	2345,946
vessel_categoryMerchant >=100GT	2153,688
sea_state2	2016,421
wind_force3	1789,818
occurrence_locationPort approach	1599,949
materialPlastic	1548,209
natural_light3	1464,074

Izvor: izradio autor

Vrijednosti važnosti predstavljaju apsolutne vrijednosti težina kodiranih značajki dobivene iz pomoćnog linearnog modela potpornih vektora. Nisu normalizirane, ne izražavaju postotke ni vjerojatnosti te nemaju unaprijed određenu gornju granicu. Stoga vrijednost 4557,332 nema samostalno značenje, nego se tumači usporedbom s ostalim vrijednostima unutar istog prikaza. Veća vrijednost označava veću apsolutnu težinu značajke u korištenom izračunu.



Slika 13. Najvažnije varijable u klasifikaciji pomorskih nezgoda kod modela potpornih vektora

Izvor: izradio autor

Primjenom modela potpornih vektora ostvarena je ukupna točnost klasifikacije od 0,70763, uz makroprosječnu preciznost od 0,65891, osjetljivost od 0,66335 i F1-mjeru od 0,65636. Navedeni rezultati opisuju uspješnost razlikovanja sedam klasa *glavnih događaja* u analiziranom testnom skupu. Dopunska analiza težina značajki izdvojila je okolišne, tehničke, administrativne i operativne značajke povezane s klasifikacijom događaja. Takvi nalazi mogu usmjeriti daljnje razmatranje čimbenika rizika i mogućih preventivnih mjera u obalnoj plovidbi.

Slika 14. prikazuje utjecaj varijabli na vjerojatnost pomorskih nezgoda primjenom metode potpornih vektora, pri čemu su prikazane samo varijable s vrijednostima većim od 0,66666. Prikazane vrijednosti predstavljaju procijenjenu vjerojatnost pripadnosti pojedine varijable određenoj klasi nezgode, dobivenu primjenom modela potpornih vektora. Ograničavanjem prikaza na vrijednosti iznad navedenog praga izdvojeni su najznačajniji čimbenici s izraženim utjecajem na klasifikaciju, čime se smanjuje utjecaj slabijih i statistički manje relevantnih varijabli.

Za klasu 1 – *Prevrtanje / Naginjanje*, ističe se varijabla *godina prijave* (razina 3) (0,72715), što upućuje na vremensku komponentu kao značajan čimbenik u pojavnosti ove vrste nezgoda. Ona nije izravan čimbenik već njena važnost proizlazi iz činjenice da algoritam prepoznaje obrasce u podacima, pri čemu *godina prijave* djeluje kao proxy varijabla koja može odražavati promjene u drugim čimbenicima, poput promjena u intenzitetu pomorskog prometa, regulatornom okviru ili strukturi promatranog skupa podataka.

Za klasu 2 – *Sudar*, dominantne su varijable poput *mjesec (prosinac)* (0,91289), *područje plovidbe – Irska* (0,91289), *područje plovidbe – Quebec* (0,85918), *segment putovanja – dvosmjerna ruta* (0,82553), *stanje mora (razina 9)* i *jačina vjetra (razina 10)* (0,81026), uz izražen utjecaj *zastave broda* i *države izvještavanja*. Ovi rezultati su ukazali na snažan utjecaj kombinacije okolišnih, prostornih i operativnih čimbenika.

Za klasu 3 – *Udaranje u nepomični objekt*, najizraženiji utjecaj imaju *područje plovidbe – Poljska* (0,95775) i *područje plovidbe – Njemačka* (0,95579), uz značajnu ulogu *stanja mora (razina 5)*, *vidljivosti (razina 3)* i *mjeseca (rujan)* (0,83826), kao i tehničkih obilježja broda.

Za klasu 4 – *Nasukavanje*, izdvajaju se *mjesec (veljača)*, *stanje mora (razina 2)*, *vidljivost (razina 2)* i *lokacija – lučko područje* (0,83363), uz dodatni utjecaj *tipa broda*, *posade* i *segmenta putovanja*, što upućuje na kombinaciju navigacijskih i operativnih uvjeta.

Za klasu 5 – *Oštećenje / Gubitak opreme*, dominantan utjecaj ima *zastava broda* (Antigva i Barbuda, Litva, Liberija) s vrijednostima većim od 0,86, što može ukazivati na razlike u standardima upravljanja, održavanja ili izvještavanja.

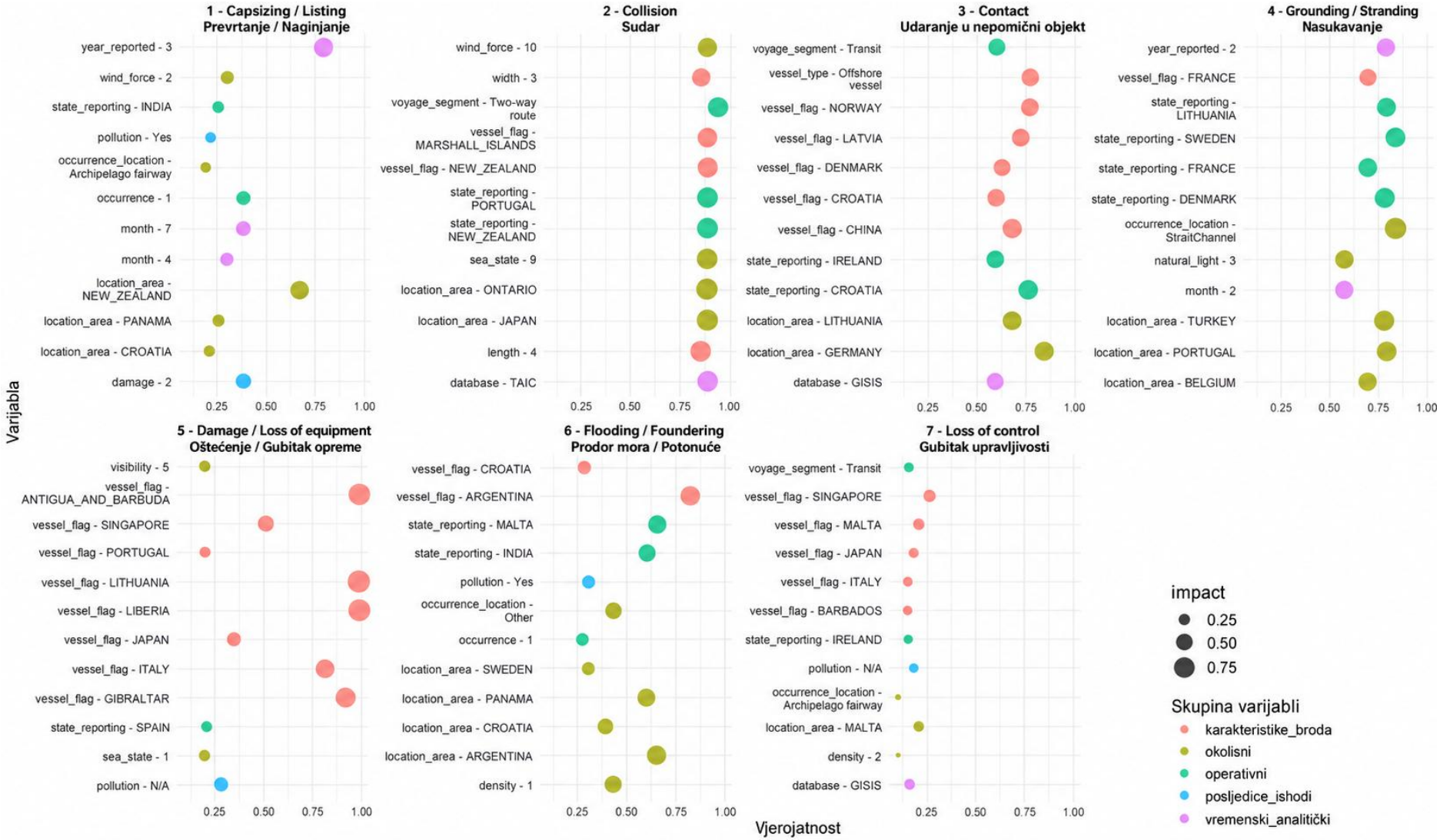
Za klasu 6 – *Prodor mora / Potonuće*, ključni su okolišni čimbenici poput *jačine vjetra (razina 4)*, *prirodne svjetlosti* i *mjeseca (kolovoz)* (0,71472), uz izražen utjecaj *tipa* i *kategorije broda*, što potvrđuje osjetljivost ove vrste nezgoda na meteorološke i tehničke uvjete.

Za klasu 7 – *Gubitak upravljivosti*, nisu identificirane varijable s vrijednostima iznad definiranog praga, što upućuje na difuznu raspodjelu utjecaja i izostanak dominantnih čimbenika unutar promatranog modela.

Analiza je jasno pokazala da se dominantni čimbenici razlikuju među klasama nezgoda, što potvrđuje kompleksnu i višedimenzionalnu prirodu pomorskog rizika. Dok su kod sudara ključni okolišni i prostorni uvjeti, kod nasukavanja prevladavaju uvjeti plovidbe, a kod oštećenja i gubitka opreme dominantnu ulogu imaju administrativne i tehničke karakteristike broda, poput *zastave broda*. Ovakva diferencijacija ukazuje na potrebu primjene specifičnih mjera upravljanja rizikom prilagođenih pojedinom tipu nezgode.

Identificirane varijable koristile su se u daljnjoj analizi za kvantitativnu procjenu rizika, pri čemu su njihove procjenjene vjerojatnosti pojave pojedinih glavnih događaja P_i kombinirane s empirijski određenim koeficijentom posljedica C_i prema IMO FSA smjernicama. Na taj se način omogućila cjelovita i kvantitativno utemeljena procjena rizika pomorskih nezgoda.

Zaključno, SVM model je pokazao zadovoljavajuću sposobnost klasifikacije i predikcije pomorskih nezgoda, osobito kod događaja s jasnije izraženim obrascima, dok su se veća odstupanja javljala kod klasa koje dijele slične operativne i okolišne značajke. Dobiveni rezultati potvrdili su da kombinacija procijenjenih vjerojatnosti pojave nezgoda P_i i empirijskog koeficijenta posljedica C_i, interpretirana u okviru IMO FSA pristupa, omogućuje određivanje razine rizika te pouzdanu identifikaciju kritičnih čimbenika i prioritetnih područja za upravljanje sigurnošću u obalnoj plovidbi.



Slika 14. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda metodom potpornih vektora

Izvor: izradio autor

U Tablici B2. iz Priloga B prikazan je kontinuirani rizik za metodu potpornih vektora, kojim se kvantificira doprinos svake varijable ukupnom riziku na temelju procijenjenih vjerojatnosti i empirijski određenih posljedica. Ovakav pristup omogućuje detaljnu i numerički preciznu analizu strukture rizika.

Na temelju tih vrijednosti formirana je matrica rizika (Slika 15.), u kojoj su kontinuirane procjene transformirane u diskretne razine učestalosti i posljedica, čime je omogućena jasna klasifikacija rizika (mali, srednji, veliki) i njegova operativna interpretacija.

Analiza pokazuje da su najviše razine rizika prisutne kod klasa *Sudar* i *Nasukavanje*, gdje dominiraju neprihvatljive vrijednosti, dok se kod klase *Prodor mora / Potonuće* javljaju i maksimalne razine rizika. Klasa *Prevrtanje / Naginjanje* karakterizirana je kombinacijom srednjih i velikih razina rizika, dok *Udaranje u nepomični objekt* pokazuje pretežno srednji rizik. Suprotno tome, klase *Oštećenje / Gubitak opreme* i *Gubitak upravljivosti* dominantno pripadaju području niskog rizika.

Dobiveni rezultati jasno diferenciraju glavne događaje prema razini rizika te potvrđuju da su *Sudar*, *Nasukavanje* i *Prodor mora / Potonuće* najkritičniji za sigurnost plovidbe.

Varijabla	1 - Capsizing / Listing	2 - Collision	3 - Contact	4 - Grounding / Stranding	5 - Damage / Loss of equipment	6 - Flooding / Foundering	7 - Loss of control
POSLJEDICA C_j	C_i=5.406	C_i=1.345	C_i=0.373	C_i=1.025	C_i=0.069	C_i=4.938	C_i=0.251
crew	R=8	R=12	R=10	R=15	R=1	R=4	R=2
damage	R=16	R=15	R=8	R=15	R=1	R=12	R=4
database	R=8	R=15	R=8	R=15	R=1	R=8	R=4
density	R=16	R=15	R=4	R=15	R=1	R=16	R=2
gt	R=12	R=12	R=10	R=12	R=2	R=20	R=4
injuries_total	R=12	R=15	R=10	R=15	R=1	R=8	R=2
length	R=8	R=15	R=10	R=15	R=1	R=8	R=4
location_area	R=12	R=15	R=8	R=15	R=2	R=12	R=4
material	R=8	R=15	R=6	R=15	R=1	R=16	R=4
month	R=12	R=15	R=10	R=15	R=2	R=16	R=4
natural_light	R=8	R=12	R=4	R=15	R=1	R=20	R=4
occurrence	R=16	R=12	R=8	R=15	R=1	R=20	R=4
occurrence_location	R=12	R=15	R=4	R=15	R=1	R=12	R=2
pilot	R=8	R=15	R=10	R=15	R=1	R=4	R=2
pollution	R=12	R=12	R=6	R=15	R=4	R=16	R=8
sea_state	R=8	R=15	R=10	R=15	R=2	R=12	R=4
ship_age	R=16	R=15	R=10	R=15	R=1	R=8	R=2
state_reporting	R=12	R=15	R=8	R=15	R=2	R=12	R=4
vessel_category	R=8	R=15	R=4	R=15	R=1	R=16	R=4
vessel_flag	R=8	R=15	R=10	R=15	R=4	R=8	R=6
vessel_sink	R=8	R=15	R=4	R=15	R=1	R=8	R=2
vessel_type	R=8	R=15	R=10	R=15	R=2	R=12	R=4
visibility	R=8	R=15	R=10	R=15	R=1	R=4	R=4
voyage_segment	R=8	R=15	R=8	R=15	R=1	R=16	R=4
weather	R=8	R=15	R=4	R=15	R=1	R=16	R=4
width	R=8	R=15	R=10	R=15	R=1	R=8	R=2
wind_force	R=12	R=15	R=10	R=15	R=2	R=12	R=4
year_reported	R=16	R=15	R=4	R=15	R=1	R=8	R=4

Razina rizika
Mali rizik
Srednji rizik
Veliki rizik

Slika 15. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih modelom potpornih vektora i pripadajućih indeksa posljedica

Izvor: izradio autor

Zaključno, kombinacija kontinuirane procjene i diskretne klasifikacije omogućuje pouzdanu i metodološki utemeljenu procjenu rizika te jasnu interpretaciju njegove razine u kontekstu sigurnosti plovidbe.

4.3. Rezultati klasifikacije, predikcije nezgoda i procjene rizika primjenom slučajnih šuma

Prilikom izgradnje modela slučajnih šuma provedena je pažljiva optimizacija ključnog hiperparametra – broja stabala u šumi (*ntree*). Testirane su vrijednosti 120, 300, 360, 480 i 500 stabala. Za svaku od ovih vrijednosti model je treniran na trening skupu, a zatim evaluiran na validacijskom skupu kako bi se utvrdila točnost predviđanja.
Optimalni broj stabala odabran je kao onaj (*ntree*=300) koji daje najveću točnost na validacijskom skupu. U ovom slučaju to je bila vrijednost koja je osigurala najbolju ravnotežu između točnosti i sposobnosti generalizacije, odnosno smanjenja prenaučenosti ili podnaučenosti (*underfitting*) modela.

Matrica zabune za model slučajnih šuma (Tablica 47.) prikazuje točnost klasifikacije pomorskih nezgoda raspoređenih po sedam klasa događaja. Konkretno, klasa 3 – *Udaranje u nepomični objekt* imala je (36) ispravno klasificiranih uzoraka, uz manji broj pogrešaka raspoređenih prema klasama 1 – *Prevrtanje / Naginjanje*, 2 – *Sudar*, 4 – *Nasukavanje*, 5 – *Oštećenje / Gubitak opreme*, 6 – *Prodor mora / Potonuće* i 7 – *Gubitak upravljivosti*. Klasa 4 – *Nasukavanje* bilježila je (33) točne predikcije, uz nekoliko pogrešnih klasifikacija u klase 2 – *Sudar*, 3 – *Udaranje u nepomični objekt* i 5 – *Oštećenje / Gubitak opreme*. Klasa 5 – *Oštećenje / Gubitak opreme* imala je (48) točno klasificiranih uzoraka, uz minimalne pogreške prema klasama 3 – *Udaranje u nepomični objekt*, 6 – *Prodor mora / Potonuće* i 7 – *Gubitak upravljivosti*.

Klase 2 – *Sudar* i 1 – *Prevrtanje / Naginjanje* imale su veći broj pogrešaka, što upućuje na poteškoće modela u njihovu razlikovanju. Klase 6 – *Prodor mora / Potonuće* i 7 – *Gubitak upravljivosti* imale su manji broj točnih predikcija, što može biti povezano s manjim brojem uzoraka ili preklapanjem obilježja s drugim klasama.

Tablica 47. Matrica zabune za model slučajnih šuma

Stvarna klasa → / Predikcija ↓	1	2	3	4	5	6	7
1	10	0	1	1	0	0	0
2	0	41	0	12	0	0	2
3	1	4	36	1	5	2	6
4	1	6	3	33	1	2	0
5	0	0	1	1	48	1	2
6	0	1	0	0	0	8	0
7	0	0	0	0	1	0	5

Izvor: izradio autor

Ključne klasifikacijske metrike (Tablica 48.) potvrđuju robusnost modela slučajnih šuma. Ukupna točnost iznosi 0,767, dok je stopa pogreške 0,233, što ukazuje na umjerenu učinkovitost u višeklasnoj klasifikaciji. Model pokazuje dobru sposobnost prepoznavanja stvarnih pozitivnih slučajeva, s osjetljivošću 0,797, te visoku pouzdanost u isključivanju pomorskih nezgoda kroz specifičnost 0,960. Preciznost 0,716 i F1-mjera 0,735 potvrđuju uravnoteženu izvedbu modela.
Visoki pozitivni omjer vjerojatnosti (31,65) i niski negativni omjer vjerojatnosti (0,21) dodatno potvrđuju sposobnost modela da razlikuje pozitivne od negativnih slučajeva. Konačni

dijagnostički omjer 200,45 potvrđuje visoku ukupnu dijagnostičku vrijednost modela slučajnih šuma.

Ukupno, rezultati pokazuju da model slučajne šume kombinira tehničke, operativne i okolišne čimbenike za učinkovitu klasifikaciju potencijalnih nezgoda u obalnoj plovidbi. Stabilnost klasifikacije ključnih klasa i visoka točnost modela potvrđuju njegovu primjenjivost kao alata za podršku odlučivanju u operativnim sigurnosnim sustavima, gdje su pravovremeno prepoznavanje i pouzdana klasifikacija glavnih događaja od ključne važnosti.

Tablica 48. Ključne klasifikacijske metrike modela slučajnih šuma

Model	RF
Accuracy	0,76695
ErrorRate	0,23305
Sensitivity (makroprosjek)	0,79694
Specificity	0,96000
Precision (makroprosjek)	0,71554
NPV	0,95882
F1 (makroprosjek)	0,73488
FDR	0,28446
FOR	0,04118
FPR	0,04000
FNR	0,20306
LR_Positive	31,65425
LR_Negative	0,21246
DOR (makroprosjek)	200,44876

Izvor: izradio autor

Tablica 49. prikazuje raspored varijabli prema njihovom doprinosu u modelu slučajnih šuma, pri čemu su varijable s najvećim utjecajem *stanje_mora* (108,679), *jačina_vjetra* (95,584) i *ozbiljnost_nezgode* (90,845), što naglašava dominantnu ulogu meteoroloških uvjeta i ozbiljnosti nezgode u klasifikaciji pomorskih nezgoda.

Sljedeće po važnosti su vremenske i tehničke varijable, uključujući *mjesec* (80,071), *oštećenje* (79,819) i *tip_broda* (61,849), dok administrativne i lokacijske varijable (*peljar* 56,879, *zastava_broda* 55,669, s*egment_plovidbe* 55,562 i *područje_nezgode* 40,721) doprinose dodatnoj preciznosti modela.

Niže rangirane varijable, poput *posada* (24,234) i *materijal* (16,861), iako manje utjecajne, doprinose finijim razlikama među klasama, osiguravajući detaljniju i stabilniju klasifikaciju nezgoda.

Ukupno, rezultati pokazuju da model slučajnih šuma kombinira meteorološke, tehničke i administrativno-lokacijske varijable za učinkovitu klasifikaciju potencijalnih nezgoda u obalnoj plovidbi, pri čemu se najvažniji čimbenici odnose na stanje mora, jačinu vjetra i ozbiljnost nezgode.

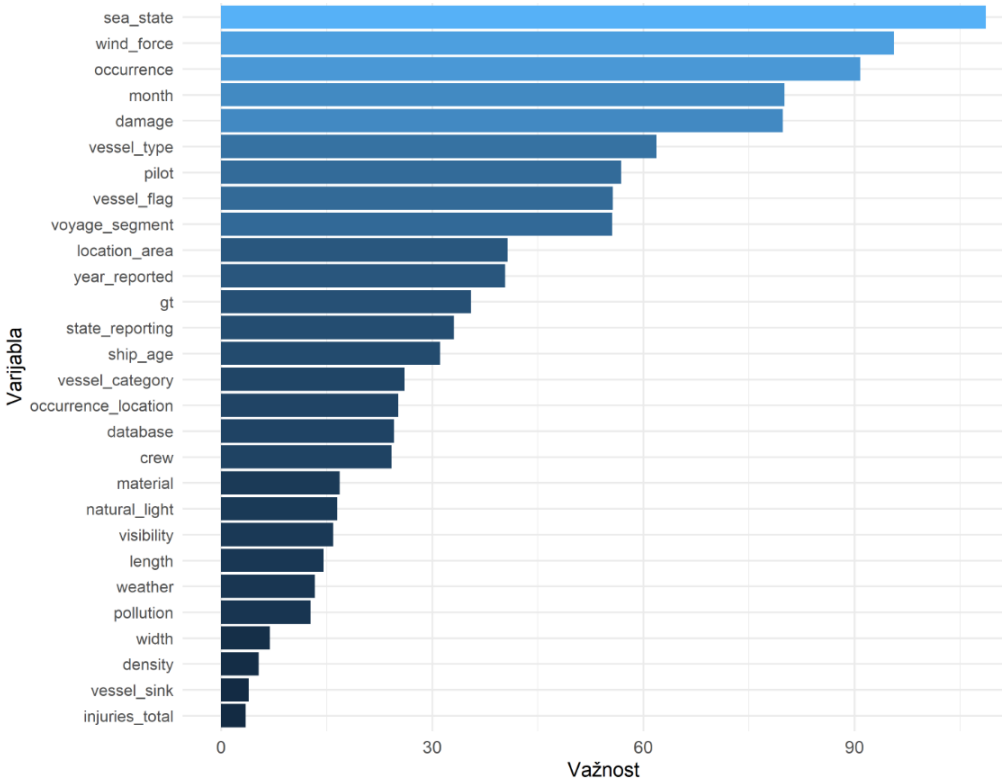
Vrijednosti važnosti u modelu slučajnih šuma, temeljene na smanjenju nečistoće čvorova, nenegativne su i nemaju unaprijed određenu gornju granicu. Veća vrijednost označava veći doprinos klasifikaciji unutar modela, a ne postotak ili statističku značajnost. Zbog različitih postupaka izračuna, njihove apsolutne vrijednosti nisu izravno usporedive s vrijednostima važnosti u modelu potpornih vektora.

Tablica 49. Varijable najsnažnijeg utjecaja u modelu slučajnih šuma

Varijabla	Važnost
sea_state	108,679
wind_force	95,584
occurrence	90,845
month	80,071
damage	79,819
vessel_type	61,849
pilot	56,879
vessel_flag	55,669
voyage_segment	55,562
location_area	40,721
year_reported	40,391
gt	35,506
state_reporting	33,081
ship_age	31,139
vessel_category	26,055
occurrence_location	25,167
database	24,612
crew	24,234
material	16,861
natural_light	16,532

Izvor: izradio autor

Slika 16. vizualno potvrđuje ove nalaze, jasno ističući dominantne čimbenike i olakšavajući identifikaciju varijabli s najvećim utjecajem na klasifikaciju. Prikaz omogućuje jednostavno uočavanje kombinacije meteoroloških, tehničkih, administrativnih i operativnih čimbenika koji zajedno oblikuju složene obrasce nezgoda i procjene rizika u obalnoj plovidbi.

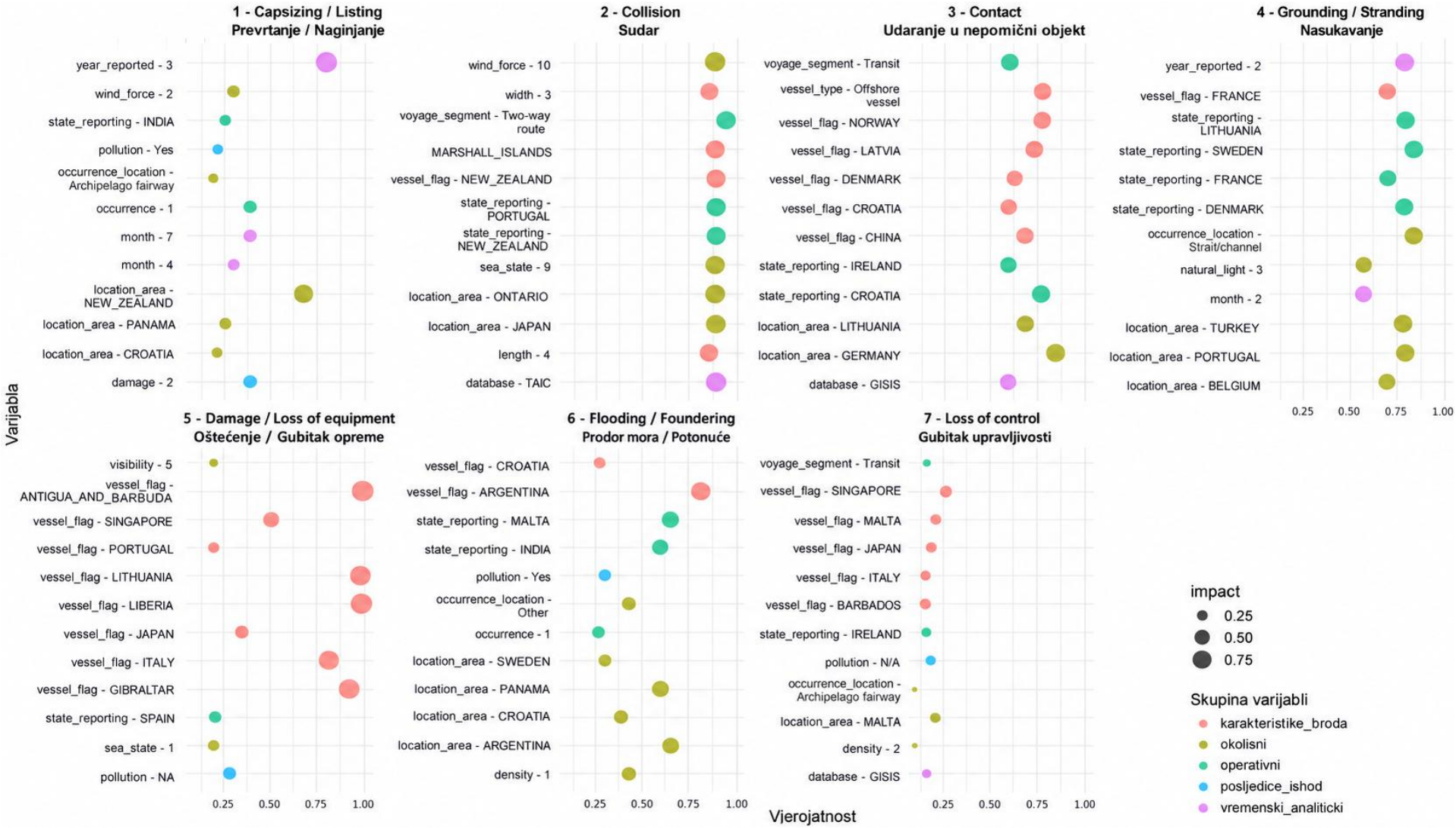


Slika 16. Najvažnije varijable u klasifikaciji pomorskih nezgoda kod modela slučajnih šuma

Izvor: izradio autor

U Tablici B3 iz Priloga B prikazane su procijenjene vjerojatnosti P_i dobivene metodom slučajnih šuma kao udio stabala koja predviđaju pojedini događaj, koje se dalje koriste u izračunu kontinuiranog rizika $R_i=P_i\times C_i$ i $R=F\times C$ te njegovoj diskretizaciji u matrici rizika radi interpretacije razine rizika.

Slika 17. prikazuje utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda primjenom metode slučajnih šuma. Model slučajne šume najveće vjerojatnosti povezuje s klasama *Sudar*, *Nasukavanje* i *Udaranje u nepomični objekt*, dok su za klase *Oštećenje / Gubitak opreme* i *Gubitak upravljivosti* vjerojatnosti znatno niže. Među najizraženijim čimbenicima izdvajaju se varijable iz skupina *okolišnih čimbenika*, *karakteristika broda* i *operativnih čimbenika*, posebno *lokacija nezgode*, *stanje mora*, *vidljivost*, *snaga vjetra*, *zastava broda*, *vrsta broda* i *segment putovanja*. Dobiveni rezultati pokazuju da metoda slučajnih šuma prepoznaje povezanost između vjerojatnosti pojave pomorskih nezgoda i istodobnog djelovanja plovidbenog okruženja, obilježja broda i operativnih uvjeta, pri čemu se najviše vrijednosti procijenjenih vjerojatnosti javljaju kod događaja povezanih sa sudarom i nasukavanjem



Slika 17. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda metodom slučajnih šuma

Izvor: izradio autor

Slika 18. pokazuje da model slučajne šume najveće procjenjene vjerojatnosti povezuje s klasama *Sudar* i *Nasukavanje*. Nakon uključivanja empirijskog koeficijenta posljedica u procjenu rizika, navedene klase izdvajaju se i kao događaji s najvišim razinama rizika. Za većinu promatranih varijabli (*gustoća prometa, područje nezgode, stanje mora, vidljivost, snaga vjetra, zastava broda, vrsta broda i segment putovanja*) utvrđeno je područje velikog rizika, dok su za klasu *Oštećenje / gubitak opreme* vrijednosti dominantno u području malog rizika, a za *Gubitak upravljivosti* uglavnom u području malog do srednjeg rizika. Rezultati ukazuju da kombinacija *okolišnih čimbenika, karakteristika broda i operativnih čimbenika* značajno utječe na obrasce vjerojatnosti pojave pojedinih vrsta nezgoda.

	1 - Capsizing / Listing	2 - Collision	3 - Contact	4 - Grounding / Stranding	5 - Damage / Loss of equipment	6 - Flooding / Foundering	7 - Loss of control
POSLJEDICA C_j	C_i=5.406	C_i=1.345	C_i=0.373	C_i=1.025	C_i=0.069	C_i=4.938	C_i=0.251
crew	R=8	R=15	R=8	R=15	R=2	R=8	R=4
damage	R=16	R=15	R=8	R=15	R=2	R=12	R=4
database	R=8	R=15	R=8	R=15	R=1	R=8	R=4
density	R=12	R=15	R=6	R=15	R=2	R=16	R=4
gt	R=12	R=12	R=10	R=15	R=3	R=12	R=4
injuries_total	R=12	R=15	R=8	R=15	R=2	R=12	R=4
length	R=8	R=15	R=8	R=15	R=2	R=8	R=4
location_area	R=12	R=15	R=8	R=15	R=2	R=12	R=4
material	R=8	R=15	R=8	R=15	R=2	R=12	R=4
month	R=16	R=15	R=10	R=15	R=2	R=12	R=4
natural_light	R=12	R=15	R=8	R=15	R=2	R=12	R=4
occurrence	R=16	R=12	R=8	R=15	R=2	R=16	R=4
occurrence_location	R=12	R=15	R=6	R=15	R=2	R=12	R=4
pilot	R=8	R=15	R=8	R=15	R=2	R=8	R=4
pollution	R=12	R=12	R=8	R=15	R=4	R=16	R=6
sea_state	R=8	R=15	R=10	R=15	R=3	R=8	R=4
ship_age	R=12	R=15	R=10	R=15	R=3	R=12	R=4
state_reporting	R=8	R=15	R=8	R=15	R=2	R=12	R=4
vessel_catbgory	R=8	R=15	R=8	R=15	R=2	R=12	R=4
vessel_flag	R=8	R=15	R=10	R=12	R=5	R=12	R=6
vessel_sink	R=8	R=15	R=6	R=15	R=2	R=8	R=4
vessel_type	R=8	R=15	R=10	R=15	R=2	R=12	R=4
visibility	R=8	R=15	R=10	R=15	R=3	R=8	R=4
voyage_segment	R=8	R=15	R=8	R=15	R=2	R=8	R=4
weather	R=8	R=15	R=8	R=15	R=2	R=12	R=4
width	R=8	R=15	R=8	R=15	R=2	R=8	R=4
wind_force	R=12	R=15	R=8	R=15	R=3	R=12	R=4
year_reported	R=16	R=15	R=4	R=15	R=1	R=12	R=2

Razina rizika

Mali rizik

Srednji rizik

Veliki rizik

Slika 18. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih modelom slučajnih šuma i pripadajućih indeksa posljedica

Izvor: izradio autor

Primjena modela slučajnih šuma potvrđuje da su *Sudar*, *Nasukavanje* i *Prodor mora / potonuće* najkritičniji događaji, pri čemu se rizik formira kao rezultat međudjelovanja okolišnih, karakteristika broda i operativnih čimbenika, što omogućuje pouzdanu identifikaciju ključnih izvora rizika u plovidbi.

Model slučajnih šuma jasno demonstrira višedimenzionalnu prirodu rizika u obalnoj plovidbi, omogućujući pouzdanu identifikaciju dominantnih čimbenika i pružajući osnovu za unaprjeđenje sustava upravljanja rizicima i preventivnih mjera.

4.4. Rezultati klasifikacije, predikcije nezgoda i procjena rizika primjenom neuronskih mreža

U ovom poglavlju prikazani su rezultati treniranja i evaluacije neuronske mreže za klasifikaciju događaja koji predstavljaju sigurnosne rizike u obalnoj plovidbi. Izbor neuronske mreže temelji se na sposobnosti ovih modela da adekvatno modeliraju složene nelinearne odnose između

ulaznih varijabli i ciljnih klasa, što je ključno u višeklasnoj klasifikaciji s velikim brojem međusobno povezanih čimbenika rizika.

Optimizacija hiperparametara bila je ključna za postizanje stabilnih performansi modela. Kao glavni hiperparametri uzeti su broj neurona u skrivenom sloju (*veličina skrivenog sloja*) i regularizacijski parametar (*decay*) koji sprječava prenaučenost. Testirane su diskretne vrijednosti: 6, 7 i 8 neurona u skrivenom sloju, te 0,5; 0,6; 0,7 i 0,8 za *decay*. Primjenom mrežne pretrage (engl. *grid search*) odabrana je optimalna konfiguracija s 6 neurona i *decay* = 0,5, jer je pokazala najbolju ravnotežu između kapaciteta modela za učenje i sposobnosti generalizacije na neviđene podatke.

Tijekom treniranja uočeni su izazovi u konvergenciji zbog velikog broja ulaznih varijabli. Kako bi se stabiliziralo učenje, primijenjena je selekcija varijabli na temelju važnosti dobivene iz modela slučajnih šuma. Postupno uklanjanje manje značajnih varijabli omogućilo je učinkovitije treniranje i bolju prediktivnu sposobnost modela.

Analiza matrice zabune (Tablica 50.) pokazala je da model *neuronske mreže* relativno dobro klasificira određene klase, dok kod drugih dolazi do značajnijih pogrešaka. Klasa *Prevrtanje / Naginjanje* (1) ostvarila je 7 točnih predikcija, uz nekoliko pogrešaka prema klasama 2, 3, 4 i 6, što pokazuje djelomičnu konfuziju s drugim klasama.

Klasa *Sudar* (2) bilježila je najveći broj ispravno klasificiranih primjera (38), uz 2 pogrešne klasifikacije prema klasi 3, 12 prema klasi 4 i po 1 pogrešku prema klasama 5 i 7, što ukazuje na sličnost karakteristika s klasom *Nasukavanje* (4). Klasa *Udaranje u nepomični objekt* (3) imala je 30 točnih predikcija, dok su pogreške raspoređene prema klasama *2*, *5*, *6* i *7*, što sugerira djelomično preklapanje obrazaca. Klasa *Nasukavanje* (4) imala je 33 ispravno klasificirana primjera, uz pogreške prema klasama *2*, *3*, *1*, *6* i *7*, što potvrđuje izazove u razlikovanju sličnih nezgoda. Klasa *Oštećenje / Gubitak opreme* (5) imala je 47 točnih predikcija, dok su pogreške raspoređene prema klasama *3* i *7*. Klasa *Prodor mora / Potonuće* (6) pokazala je 7 ispravno klasificiranih primjera, uz pogreške prema klasama *1*, *2* i *4*. Klasa *Gubitak upravljivosti* (7) pokazala je najnižu stabilnost s 7 točnih klasifikacija i pogreškama prema klasama *5* i *6*.

Ukupno, matrica zabune potvrđuje da su najstabilnije klase *2*, *4* i *5*, dok su klase *6* i *7* manje stabilne. Ovi rezultati naglašavaju potrebu za dodatnim finim podešavanjem modela neuronske mreže kako bi se poboljšala točnost kod klasa s preklapajućim karakteristikama.

Tablica 50. Matrica zabune za model neuronske mreže

Stvarna klasa → / Predikcija ↓	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	7	1	1	2	0	1	0
2	0	38	2	12	1	0	1
3	2	4	30	0	3	3	2
4	1	8	5	33	1	1	1
5	0	0	3	0	47	1	4
6	2	1	0	1	0	7	0
7	0	0	0	0	3	0	7

Izvor: izradio autor

Ključne klasifikacijske metrike (Tablica 51.) potvrđuju umjerenu učinkovitost modela neuronske mreže u klasifikaciji nezgoda u obalnoj plovidbi. Ukupna *točnost* iznosi 0,716, što

znači da je približno 72 % svih nezgoda ispravno klasificirano, dok je *stopa pogreške* 0,284, što pokazuje prostor za daljnja poboljšanja, posebno kod složenih ili rijetkih događaja.
Osjetljivost modela iznosi 0,689, što ukazuje da model relativno dobro prepoznaje stvarne pozitivne slučajeve, tj. nezgode koje zaista pripadaju određenoj klasi. S druge strane, visoka *specifičnost* od 0,951 potvrđuje da model učinkovito prepoznaje negativne slučajeve, odnosno uspješno izbjegava lažno pozitivne klasifikacije.
Preciznost modela iznosi 0,656, a *F1-mjera* 0,668, što upućuje na umjerenu ravnotežu između preciznosti i osjetljivosti, usporedivu s rezultatima ostalih analiziranih modela. Vrijednosti *pozitivnog omjera vjerojatnosti* (16,91) i *negativnog omjera vjerojatnosti* (0,327) potvrđuju dobru sposobnost modela u razlikovanju slučajeva koji pripadaju promatranoj klasi od onih koji joj ne pripadaju. Konačni *dijagnostički omjer* 57,39 ukazuje na zadovoljavajuću ukupnu dijagnostičku moć modela, što znači da neuronska mreža može pouzdano doprinijeti klasifikaciji glavnih događaja u operativnim uvjetima.

Tablica 51. Ključne klasifikacijske metrike modela neuronskih mreža

Model	NN
Accuracy	0,71610
ErrorRate	0,28390
Sensitivity (makroprosjek)	0,68854
Specificity	0,95094
Precision (makroprosjek)	0,65614
NPV	0,95040
F1 (makroprosjek)	0,66822
FDR	0,34386
FOR	0,04960
FPR	0,04906
FNR	0,31146
LR_Positive	16,90558
LR_Negative	0,32733
DOR (makroprosjek)	57,38777

Izvor: izradio autor

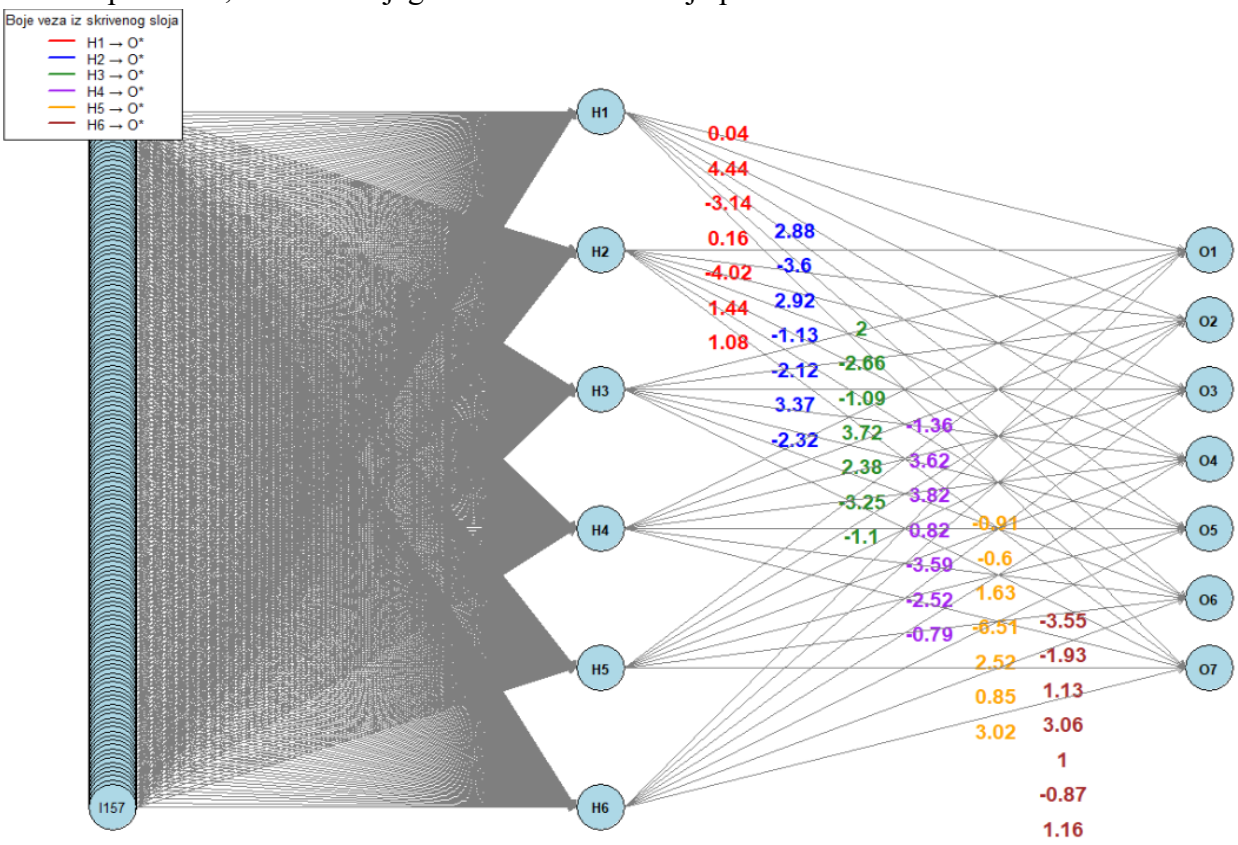
Model neuronske mreže prikazan na Slici 19. generiran je i identificiran kao optimalna konfiguracija na temelju performansi ostvarenih na validacijskom skupu podataka. Arhitektura modela definirana je s jednim skrivenim slojem koji sadrži šest neuronskih jedinica, čime je postignuta učinkovita sposobnost učenja složenih i nelinearnih odnosa uz očuvanje stabilnosti i generalizacijskih svojstava modela. Ulazni sloj obuhvaća 156 varijabli, koje su rezultat prethodno provedene redukcije dimenzionalnosti i selekcije značajki. U proces treniranja uključene su samo varijable s najvećim informacijskim doprinosom, što je omogućilo smanjenje rizika od prenaučenosti, povećanje računske učinkovitosti te poboljšanje pouzdanosti rezultata na validacijskom skupu. U strukturu ulaznog sloja uključena je bias jedinica, čija je uloga omogućavanje pomaka aktivacijskih funkcija i fleksibilnijeg prilagođavanja distribuciji podataka. Time se povećava sposobnost neuronske mreže za aproksimaciju nelinearnih odnosa, osobito u situacijama kada su ulazne varijable međusobno slične ili koncentrirane oko nižih vrijednosti. Za kontrolu složenosti modela i sprječavanje preučenja primijenjena je L2 regularizacija, pri čemu je optimalna vrijednost parametra *decay*

values utvrđena na 0,5. Ovakav pristup uvodi penalizaciju većih težina u mreži te potiče učenje općih obrazaca, čime se osigurava bolja generalizacija performansi s validacijskog na testni skup podataka.

Na temelju definiranih evaluacijskih kriterija, navedena konfiguracija, neuronska mreža s jednim skrivenim slojem, šest neuronskih jedinica, 156 ulaznih varijabli i *decay values* = 0,5, ostvarila je najbolje rezultate među svim generiranim konfiguracijama te je korištena za daljnju analizu i evaluaciju klasifikacijskih performansi.

Grafički prikaz na Slici 19. prikazuje strukturu neuronske mreže s 156 kodiranih ulaza mreže, pripadajućom bias jedinicom, jednim skrivenim slojem i sedam izlaznih neurona, koji odgovaraju definiranim klasifikacijskim kategorijama sigurnosnih događaja u obalnoj plovidbi. Zbog velikog broja ulaznih varijabli, pojedinačne težine između ulaznog i skrivenog sloja nisu prikazane kako bi se očuvala preglednost vizualnog prikaza.

Sve težine modela pohranjene su u pripadajućim Excel tablicama, što omogućuje njihovu detaljnu naknadnu analizu i dodatnu vizualizaciju prema potrebi. U samom grafičkom prikazu veze između skrivenog sloja i izlaznih neurona označene su istom bojom, pri čemu boja ne predstavlja numeričku vrijednost težina, već služi isključivo za označavanje pripadnosti istoj skupini veza, čime se izbjegava vizualno zasićenje prikaza.



Slika 19. Struktura neuronske mreže

Izvor: izradio autor

Iako su sve veze između pojedinačnog skrivenog sloja i svih izlaza prikazane istom bojom, vrijednosti težina su jasno označene brojčano. Težine mogu biti pozitivne ili negativne:

- Pozitivna težina znači da aktivacija jedinice u skrivenom sloju povećava vjerojatnost aktivacije određenog izlaznog neurona (odnosno povećava vjerojatnost pripadnosti toj klasi).
- Negativna težina ukazuje na to da aktivacija te skrivene jedinice smanjuje vjerojatnost aktivacije konkretnog izlaznog neurona, tj. da ta jedinica ima inhibicijski utjecaj na tu izlaznu klasu.

Tablica 52. prikazuje dvadeset varijabli s najvećim doprinosom u modelu neuronske mreže za klasifikaciju nezgoda u obalnoj plovidbi. Najveći doprinos daje varijabla *ozbiljnost nezgode 4* s vrijednošću važnosti 7,784. Slijedi varijabla *snaga vjetra 10* s vrijednošću 6,200, dok se na trećem mjestu nalazi *stanje mora 2* s vrijednošću 5,340.

Značajan utjecaj imaju i varijable povezane s posljedicama nezgode, među kojima se ističu *oštećenje 4* (5,291) te *ozbiljnost nezgode 3* (4,788). Administrativne karakteristike događaja također doprinose klasifikaciji, posebno *baza podataka MAIB* (4,414) i *baza podataka UEIM* (3,552), kao i varijabla *država izvjestiteljica Kina* (3,935).

Operativni kontekst plovidbe odražava varijabla *segment plovidbe dolazak* s važnošću 4,075, dok je vremenska dimenzija uključena putem *godina izvještavanja 5* s vrijednošću 4,253. Tehničke karakteristike broda imaju vidljiv doprinos kroz varijable *vrsta broda RO-RO putnički* (4,367) i *vrsta broda tegljač* (3,620).

Meteorološki čimbenici dodatno se očituju kroz varijable *snaga vjetra 5* (4,265) i *stanje mora 7* (3,326). Geografske značajke zastupljene su varijablama *područje Hong Kong* (4,124), *područje Turska* (3,439) i *područje Italija* (3,317).

Zastava broda i administrativni kontekst također doprinose modelu, što je vidljivo kod varijable *zastava Finska* s važnošću 3,618. Konačno, varijabla *ozbiljnost nezgode 2* s vrijednošću 3,915 upućuje na doprinos obilježja same nezgode klasifikaciji. Ukupno gledano, raspodjela važnosti u Tablici 52. pokazuje da model *neuronske mreže* integrira obilježja događaja, meteorološke uvjete, tehničke značajke broda te geografsko-administrativni kontekst, što upućuje na višedimenzionalnu prirodu rizika u obalnoj plovidbi. Vrijednosti važnosti u Tablici 52. kreću se od 3,317 do 7,784. Navedene granice predstavljaju raspon izdvojenih značajki, a ne teorijske granice ljestvice važnosti. Veća vrijednost označava veću važnost, ali sama po sebi ne označava statističku značajnost niti smjer utjecaja. Zbog različitih postupaka izračuna i ljestvica, apsolutne vrijednosti nisu izravno usporedive s vrijednostima važnosti u modelima slučajnih šuma i potpornih vektora, pa se usporedba prvenstveno temelji na poretku značajki.

Tablica 52. Varijable najsnažnijeg utjecaja u modelu neuronske mreže

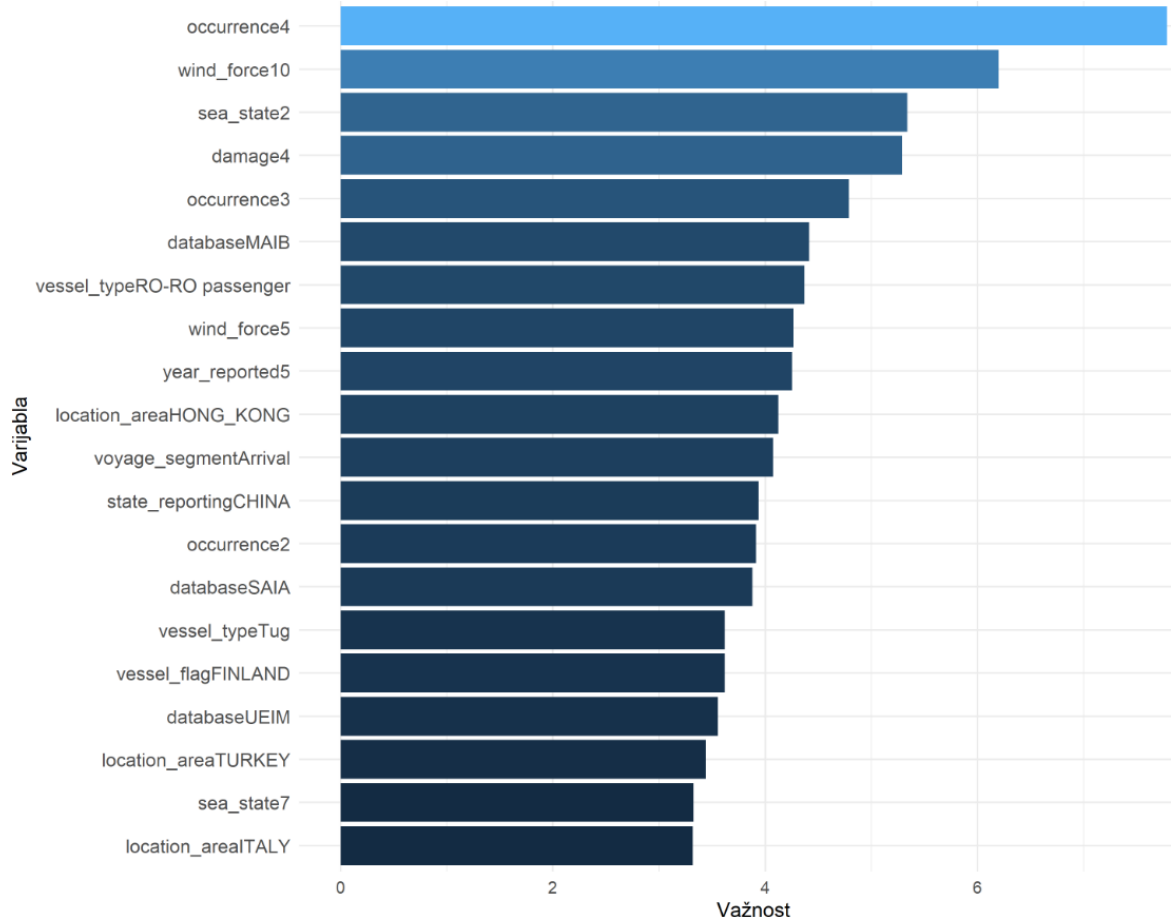
Varijabla	Važnost
occurrence4	7,784
wind_force10	6,200
sea_state2	5,340
damage4	5,291
occurrence3	4,788
databaseMAIB	4,414
vessel_typeRO-RO passenger	4,367
wind_force5	4,265
year_reported5	4,253
location_areaHONG_KONG	4,124

voyage_segmentArrival	4,075
state_reportingCHINA	3,935
occurrence2	3,915
databaseSAIA	3,878
vessel_typeTug	3,620
vessel_flagFINLAND	3,618
databaseUEIM	3,552
location_areaTURKEY	3,439
sea_state7	3,326
location_areaITALY	3,317

Izvor: izradio autor

Slika 20. grafički prikazuje relativnu važnost varijabli unutar modela neuronske mreže, gdje dominiraju varijable vezane uz ozbiljnost nezgoda i okolišne čimbenike, ali su jasno vidljive posljedice i vremensko analitički čimbenici.

Zaključno, rezultati jasno pokazuju da model neuronske mreže oslanja svoju klasifikaciju na složenu mrežu čimbenika, gdje su ključni utjecaji ozbiljnost nezgoda i vremenski uvjeti, ali ne zanemaruje se ni uloga administrativnih i tehničkih parametara. Ova multidimenzionalnost varijabli omogućava modelu da bolje razumije složene sigurnosne rizike u obalnoj plovidbi, ali istovremeno ukazuje na potrebu daljnjeg istraživanja i eventualnog uključivanja dodatnih podataka za poboljšanje točnosti i pouzdanosti modela.

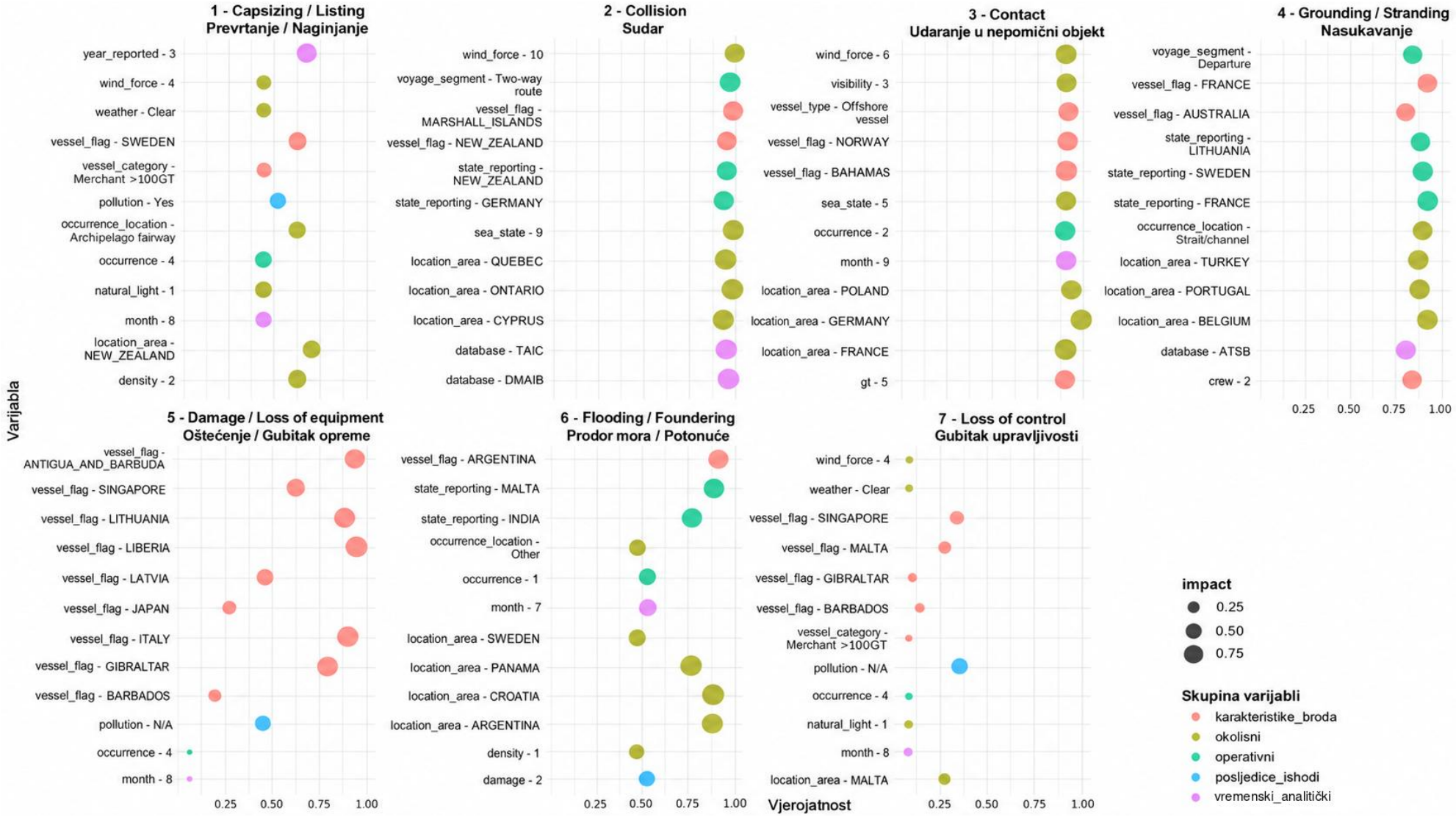


Slika 20. Najvažnije varijable u klasifikaciji pomorskih nezgoda kod modela neuronskih mreža

Izvor: izradio autor

Slika 21. prikazuje utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda primjenom modela neuronskih mreža. Za razliku od metode slučajnih šuma, neuronske mreže pokazuju izraženu koncentraciju vjerojatnosti u klasama *Nasukavanje*, *Sudar* i *Udaranje u nepomični objekt*, pri čemu se klasa *Udaranje u nepomični objekt* znatno češće javlja kao dominantna nego kod modela slučajnih šuma. Vjerojatnosti za klase *Oštećenje / Gubitak opreme* i *Gubitak upravljivosti* ostaju vrlo niske i rijetko imaju značajniji utjecaj.
Među najvažnijim čimbenicima ističu se varijable vezane uz *lokaciju nezgode*, *područje plovidbe*, *gustoću prometa* i *segment putovanja*, koje pokazuju visoke vrijednosti vjerojatnosti, osobito za klasu *Nasukavanje*. Okolišni uvjeti, uključujući *stanje mora*, *vidljivost* i *snagu vjetra*, snažno su povezani s klasama *Udaranje u nepomični objekt* i *Nasukavanje*, pri čemu u pojedinim kategorijama upravo *Udaranje* ostvaruje najviše vrijednosti vjerojatnosti. Varijable koje opisuju karakteristike broda, kao što su *zastava broda*, *vrsta broda*, *duljina* i *starost*, dodatno diferenciraju raspodjelu vjerojatnosti i potvrđuju prisutnost izraženih nelinearnih odnosa.

Dobiveni rezultati pokazuju da model neuronskih mreža procjenjuje vjerojatnost pojave glavnih događaja na temelju složenih interakcija između okolišnih, prostornih i operativnih čimbenika, pri čemu se najviše vrijednosti procijenjenih vjerojatnosti sustavno vežu uz događaje *Nasukavanje* i *Udaranje u nepomični objekt*, uz značajan doprinos klase *Sudar*.
U Tablici B4. iz Priloga B prikazan je kontinuirani rizik dobiven primjenom modela neuronskih mreža, kojim se kvantificira doprinos pojedinih varijabli ukupnom riziku na temelju raspodjele vjerojatnosti po sedam klasa glavnih događaja i pripadajućih težina posljedica. Ovakav pristup omogućuje preciznije zahvaćanje nelinearnih odnosa među varijablama te realističniju procjenu strukture rizika u odnosu na klasične metode.



Slika 21. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda metodom neuronskih mreža

Izvor: izradio autor

Na temelju kontinuiranih vrijednosti formirana je matrica rizika (Slika 22.), pri čemu su procijenjene vjerojatnosti pojave glavnih događaja kategorizirane u razine učestalosti te kombinirane s odgovarajućim razinama posljedica. Na taj način je omogućena jasna klasifikaciju događaja prema razinama rizika na mali, srednji i veliki, čime je omogućena njihova daljnja operativna primjena.

Rezultati pokazuju izrazitu dominaciju klasa *Nasukavanje* i *Sudar*, koje u velikom broju analiziranih varijabli ostvaruju visoke razine rizika kao rezultat kombinacije povećanih vjerojatnosti pojave i pripadajućih posljedica. *Nasukavanj*e je posebno naglašeno kod prostornih i navigacijskih obilježja (npr. *lokacija, gustoća prometa, područje plovidbe*), dok je *Sudar* izražen kod administrativnih i operativnih varijabli (*država izvjestiteljica, zastava broda, tip i kategorija plovila*). Klasa *Udaranje u nepomični ob*jekt zauzima srednje do visoke razine rizika, osobito u specifičnim uvjetima plovidbe i kod određenih tipova brodova, što ukazuje na njezinu operativnu relevantnost.

Klasa *Prodor mora / Potonuće* pokazuje selektivno visoke razine rizika povezane s ozbiljnošću posljedica, stanjem mora i određenim geografskim područjima, čime potvrđuje visok potencijal posljedica tih događaja. Suprotno tome, *Oštećenje / Gubitak opreme* i *Gubitak upravljivosti* u većini slučajeva ostaju u zoni niskog rizika, uz ograničene iznimke.

Dobiveni rezultati jasno izdvajaju *Sudar* i *Nasukavanje* kao dominantne izvore rizika, uz značajan doprinos *Prodora mora / Potonuća*, čime se potvrđuje konzistentnost s obrascima sigurnosno najkritičnijih događaja u plovidbi.

	1 - Capsizing / Listing	2 - Collision	3 - Contact	4 - Grounding / Stranding	5 - Damage / Loss of equipment	6 - Flooding / Foundering	7 - Loss of control
POSLJEDICA C_j	C_i=5.406	C_i=1.345	C_i=0.373	C_i=1.025	C_i=0.069	C_i=4.938	C_i=0.251
crew	R=8	R=15	R=8	R=15	R=1	R=4	R=2
damage	R=12	R=15	R=10	R=15	R=1	R=16	R=2
database	R=4	R=15	R=6	R=15	R=1	R=8	R=4
density	R=16	R=15	R=4	R=15	R=1	R=16	R=2
gt	R=16	R=12	R=10	R=15	R=2	R=16	R=4
injuries_total	R=16	R=15	R=8	R=15	R=1	R=12	R=4
length	R=8	R=15	R=10	R=12	R=1	R=8	R=4
location_area	R=12	R=15	R=8	R=12	R=1	R=16	R=4
material	R=16	R=15	R=6	R=15	R=2	R=12	R=4
month	R=16	R=15	R=10	R=15	R=1	R=12	R=4
natural_light	R=16	R=12	R=4	R=15	R=2	R=16	R=4
occurrence	R=20	R=9	R=10	R=15	R=2	R=20	R=4
occurrence_location	R=16	R=12	R=4	R=15	R=1	R=12	R=2
pilot	R=8	R=12	R=10	R=15	R=1	R=4	R=2
pollution	R=16	R=12	R=4	R=15	R=4	R=12	R=8
sea_state	R=12	R=15	R=10	R=15	R=1	R=8	R=4
ship_age	R=16	R=15	R=10	R=15	R=1	R=8	R=4
state_reporting	R=8	R=15	R=6	R=15	R=1	R=12	R=4
vessel_category	R=12	R=15	R=4	R=15	R=2	R=12	R=4
vessel_flag	R=8	R=15	R=10	R=15	R=4	R=8	R=4
vessel_snk	R=8	R=15	R=2	R=15	R=1	R=4	R=2
vessel_type	R=12	R=15	R=10	R=15	R=1	R=12	R=4
visibility	R=8	R=12	R=10	R=15	R=1	R=4	R=4
voyage_segment	R=12	R=15	R=6	R=15	R=1	R=12	R=4
weather	R=16	R=15	R=4	R=15	R=2	R=12	R=4
width	R=8	R=15	R=10	R=15	R=1	R=8	R=4
wind_force	R=12	R=15	R=10	R=15	R=1	R=12	R=4
year_reported	R=16	R=15	R=4	R=15	R=1	R=12	R=2

Razina rizika
Mali rizik
Srednji rizik
Veliki rizik

Slika 22. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih modelom neuronskih mreža i pripadajućih indeksa posljedica

Izvor: izradio autor

Zaključno, model neuronskih mreža omogućuje dublje razumijevanje strukture rizika kroz identifikaciju dominantnih obrazaca i međuzavisnosti, čime pruža pouzdanu osnovu za donošenje odluka u upravljanju sigurnošću plovidbe. Model neuronske mreže pokazuje

umjerene performanse u klasifikaciji i predikciji nezgoda u obalnoj plovidbi. Ukupna *točnost* iznosi 0,712, dok je *specifičnost* 0,950, što znači da model pouzdano prepoznaje negativne slučajeve. *Osjetljivost* od 0,671 i *F1-mjera* od 0,662 ukazuju na umjerenu sposobnost prepoznavanja stvarnih pozitivnih događaja. Međutim, zapažen je veći broj pogrešaka kod pojedinih klasa, posebno kod rijetkih ili preklapajućih događaja, poput *Prevrtanje / Naginjanje* i *Gubitak upravljivosti*. Ovi rezultati naglašavaju važnost dodatne optimizacije modela i pažljivog odabira varijabli za stabilno učenje u višeklasnim uvjetima.

4.5. Usporedni prikaz rezultata klasifikacije, predikcije pomorskih nezgoda i procjene rizika primjenom potpornih vektora, slučajnih šuma i neuronskih mreža

Radi cjelovitog vrednovanja primijenjenih modela provedena je usporedna analiza njihove uspješnosti u klasifikaciji glavnih događaja pomorskih nezgoda te procjene vjerojatnosti pojave pojedinih klasa, koje predstavljaju osnovu za naknadnu kvantifikaciju razine rizika. Usporedba obuhvaća broj pravilno klasificiranih slučajeva po glavnim događajima te skup standardnih mjera uspješnosti klasifikacije, čime se omogućuje pouzdana procjena njihove primjenjivosti u analizi sigurnosti plovidbe.

U Tablici 53. prikazan je broj pravilno klasificiranih slučajeva za svaku klasu glavnih događaja: *Prevrtanje / Naginjanje*, *Sudar*, *Udaranje u nepomični objekt*, *Nasukavanje*, *Oštećenje / Gubitak opreme*, *Prodor mora / Potonuće* i *Gubitak upravljivosti*. Rezultati su pokazali da model slučajnih šuma ostvaruje najveći ukupan broj točnih klasifikacija (181), dok su potporni vektori i neuronske mreže postigle vrlo slične rezultate (167 i 169).

Najveće razlike među modelima uočene su kod učestalijih i složenijih događaja. Kod klase *Sudar* (2) model slučajnih šuma ostvario je 41 pravilnu klasifikaciju, dok su neuronske mreže postigle 38, a potporni vektori 32. Sličan obrazac vidljiv je i kod klase *Udaranje u nepomični objekt* (3), gdje su slučajne šume imale 36 točnih klasifikacija, potporni vektori 34, a neuronske mreže 30. Kod klase *Nasukavanje* (4) svi modeli su ostvarili jednak rezultat (33), što upućuje na stabilnu prepoznatljivost ovog događaja neovisno o korištenoj metodi. Klasa *Oštećenje / Gubitak opreme* (5) je pokazala visoku točnost kod svih modela, uz minimalne razlike. Za rjeđe događaje, poput *Prodor mora / Potonuće* (6) i *Gubitak upravljivosti* (7), rezultati su niži i varijabilniji, pri čemu su neuronske mreže postigle najbolji rezultat kod *Gubitka upravljivosti* (7), a potporni vektori kod *Prodora mora / Potonuća* (6).

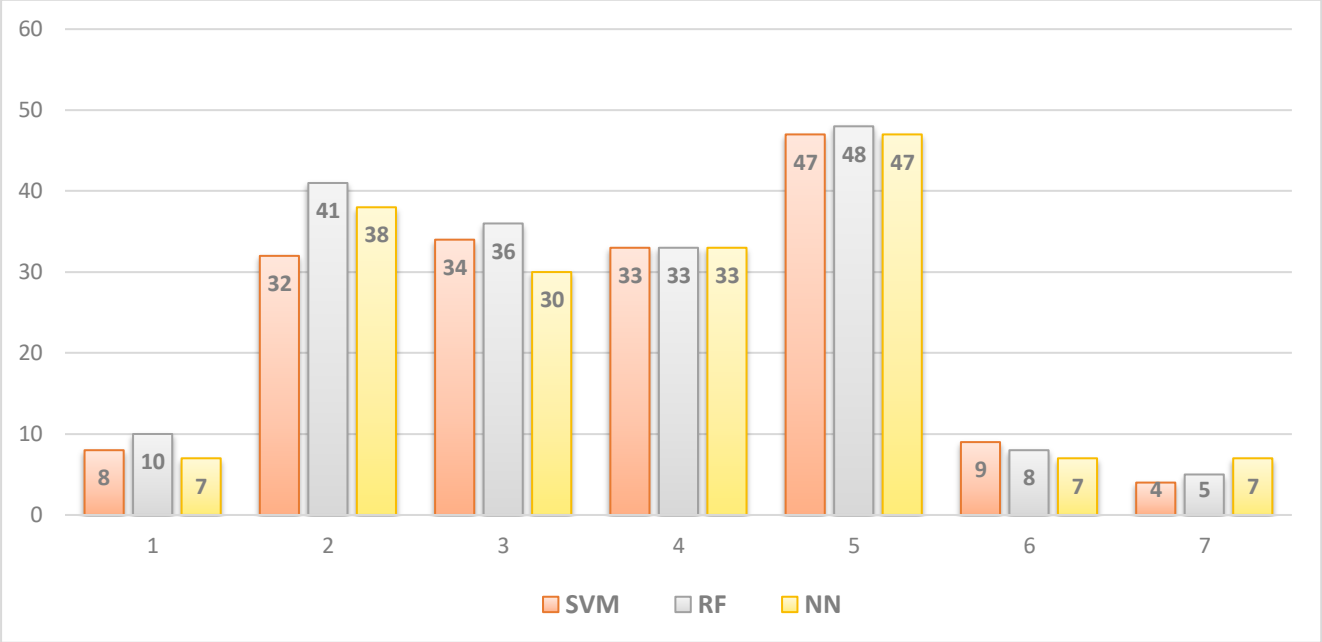
Tablica 53. Pravilno klasificirani slučajevi na testnom skupu

Klasa	SVM	RF	NN
1	8	10	7
2	32	41	38
3	34	36	30
4	33	33	33
5	47	48	47
6	9	8	7
7	4	5	7
Ukupno	167	181	169

Izvor: izradio autor

Grafički prikaz na Slici 23. potvrđuje ove odnose te jasno ističe dominaciju modela slučajnih šuma u većini klasa, osobito kod *Sudar* (2) i *Udaranje u nepomični objekt* (3).

Potporni vektori i neuronske mreže su pokazali ujednačeniju, ali nižu razinu točnosti, uz manje oscilacije između pojedinih klasa.



Slika 23. Usporedni prikaz točno klasificiranih slučajeva po glavnim događajima i modelima (SVM, RF, i NN)

Izvor: izradio autor

Detaljnija analiza uspješnosti prikazana je u Tablici 54., gdje su dani ključni pokazatelji klasifikacije. Najveću ukupnu točnost ostvario je model slučajnih šuma (0,76695), dok su neuronske mreže (0,71610) i potporni vektori (0,70763) postigli nešto slabije rezultate. Slučajne šume također su imale najvišu osjetljivost (0,79694), specifičnost (0,96000), preciznost (0,71554) i F1 mjeru (0,73488), što ukazuje na njihovu uravnoteženu i pouzdanu sposobnost klasifikacije. Neuronske mreže u većini pokazatelja su ostvarile bolje rezultate od potpornih vektora, osobito u osjetljivosti i F1 mjeri, ali i dalje su zaostale za modelom slučajnih šuma.

Vrijednosti dijagnostičkog omjera dodatno je potvrdio ove nalaze, pri čemu su slučajne šume ostvarile znatno višu vrijednost (200,44876) u odnosu na potporne vektore i neuronske mreže, što je ukazalo na njihovu veću diskriminacijsku sposobnost.

Tablica 54. Analiza metrika klasifikacije

Model	SVM	RF	NN
Accuracy	0,70763	0,76695	0,71610
ErrorRate	0,29237	0,23305	0,28390
Sensitivity (makroprosjek)	0,66335	0,79694	0,68854
Specificity	0,94971	0,96000	0,95094
Precision (makroprosjek)	0,65891	0,71554	0,65614
NPV	0,94915	0,95882	0,95040
F1 (makroprosjek)	0,65636	0,73488	0,66822
FDR	0,34109	0,28446	0,34386
FOR	0,05085	0,04118	0,04960
FPR	0,05029	0,04000	0,04906

FNR	0,33665	0,20306	0,31146
LR_Positive	19,18106	31,65425	16,90558
LR_Negative	0,35454	0,21246	0,32733
DOR (makroprosjek)	62,69772	200,44876	57,38777

Izvor: izradio autor

Rezultati iz Tablice 53., Slike 23. i Tablice 54. jasno pokazuju da model slučajnih šuma ostvaruje najbolju ukupnu učinkovitost i stabilnost u klasifikaciji glavnih događaja pomorskih nezgoda. Potporni vektori i neuronske mreže daju usporedive rezultate, pri čemu neuronske mreže pokazuju blagu prednost u odnosu na potporne vektore, ali bez dosezanja razine učinkovitosti modela slučajnih šuma. Modeli potpornih vektora i neuronskih mreža pokazuju konzistentne, ali slabije performanse u odnosu na model slučajnih šuma, što upućuje na potrebu njihovog dodatnog podešavanja ili integracije u naprednije pristupe, poput ansambla ili slojevitih modela, koji će se detaljnije razmotriti u nastavku.
Usporedni prikaz vjerojatnosti točno klasificiranih slučajeva po glavnim događajima i modelima pokazuje da sva tri pristupa dosljedno izdvajaju *Sudar* i *Nasukavanje* kao sigurnosno najkritičnije događaje, uz razlike u raspodjeli vjerojatnosti i naglašavanju pojedinih klasa.

Rezultati modela potpornih vektora na Slici 14. ukazuju na izražen utjecaj okolišnih, prostornih i operativnih varijabli, pri čemu su najviše vrijednosti vjerojatnosti vezane uz *Sudar* i *Nasukavanje*, dok se *Udaranje u nepomični objekt* javlja u specifičnim uvjetima plovidbe. Kontinuirani rizik u Tablici B2. i matrica rizika na Slici 15. potvrđuju dominaciju visokih razina rizika kod *Sudar* i *Nasukavanje*, uz povremene maksimalne vrijednosti za *Prodor mora / Potonuće*, dok *Oštećenje / Gubitak opreme* i *Gubitak upravljivosti* ostaju pretežno u području niskog rizika.

Model slučajnih šuma na Slici 17. pokazuje uravnoteženiju raspodjelu vjerojatnosti te jasno ističe klase *Sudar*, *Nasukavanje* i *Udaranje u nepomični objekt* kao dominantne. Vjerojatnosti su snažno povezane s kombinacijom okolišnih uvjeta, obilježja broda i operativnih čimbenika, čime se dodatno potvrđuje složenost i višedimenzionalnost pomorskog rizika.

Vjerojatnosti modela neuronskih mreža prikazane u Tablici B4., zajedno s grafičkim prikazom na Slici 21., pokazuju izraženu koncentraciju u klasama *Nasukavanje*, *Sudar* i *Udaranje u nepomični objekt*, pri čemu *Udaranje u nepomični objekt* ima naglašeniju ulogu nego kod ostalih modela. Kontinuirani rizik u Tablici B4. i matrica rizika na Slici 22. potvrđuju da se *Nasukavanje* i *Sudar* u velikom broju varijabli nalaze u području visokog rizika. *Udaranje u nepomični objekt* zauzima srednje do visoke razine rizika, dok *Prodor mora / Potonuće* pokazuje selektivno visoke vrijednosti povezane s okolišnim i operativnim uvjetima. *Oštećenje / Gubitak opreme* i *Gubitak upravljivosti* ostaju uglavnom u zoni niskog rizika.

Rezultati prikazani na Slikama 14., 15., 17., 18., 21. i 22. te u Tablicama B2., B3. i B4. potvrđuju visoku razinu konzistentnosti među modelima u identifikaciji ključnih izvora rizika. *Sudar* i *Nasukavanje* izdvajaju se kao dominantni događaji, dok neuronske mreže dodatno naglašavaju važnost *Udaranja u nepomični objekt*, čime pružaju dublji uvid u složene i nelinearne obrasce pomorskih nezgoda.
Rezultati usporedne analize procijenjenih vjerojatnosti pojave glavnih događaja i pripadajućih razina rizika potvrđuju visoku razinu konzistentnosti među modelima u identifikaciji ključnih

događaja, pri čemu se *Sudar* i *Nasukavanje* sustavno izdvajaju kao dominantni izvori rizika, uz dodatno naglašenu ulogu *Udaranja u nepomični objekt* kod modela neuronskih mreža. Unatoč razlikama u raspodjeli procijenjenih vjerojatnosti glavnih događaja i interpretaciji utjecajnih čimbenika, svi modeli pružaju komplementaran uvid u strukturu pomorskog rizika.
Uočene razlike u raspodjeli procijenjenih vjerojatnosti pojave glavnih događaja između modela potpornih vektora, slučajnih šuma i neuronskih mreža ukazuju na njihovu komplementarnost, ali i na ograničenja pojedinačnih pristupa u potpunom obuhvatu složenih obrazaca pomorskih nezgoda. Stoga se u nastavku primjenjuju ansambl i slojeviti modeli s ciljem integracije njihovih prednosti i poboljšanja točnosti, stabilnosti i robusnosti procjene. Nakon toga slijedi usporedni prikaz svih modela, uključujući i proširene pristupe, dok se konačna validacija provodi na objedinjeno definiranom modelskom okviru radi pouzdane procjene njegove generalizacijske sposobnosti i primjenjivosti u okviru procjene pomorskog rizika.

4.6. Ansambl model

U svrhu poboljšanja stabilnosti i pouzdanosti klasifikacije primijenjen je ansambl model koji kombinira predikcije modela potpornih vektora, slučajnih šuma i neuronskih mreža, čime se postiže uravnoteženija procjena i smanjuje utjecaj pojedinačnih pogrešaka.

Rezultati prikazani u Tablici 55. pokazuju da ansambl model zadržava visoku točnost kod dominantnih klasa 2 - *Sudar* (38), 3 - *Udaranje u nepomični objekt* (34) i 4 - *Nasukavanje* (33), kao i vrlo dobar rezultat za 5 - *Oštećenje / Gubitak opreme* (48). Kod klasa 1 - *Prevrtanje / Naginjanje* i 6 - *Prodor mora / Potonuće* postiže umjerenu, ali stabilnu točnost (po 8), dok 7 - *Gubitak upravljivosti* ostaje najmanje pouzdano klasificirana klasa (5), što je u skladu s ranijim modelima.
Pogreške su najizraženije između klasa 2 - *Sudar* i 4 - *Nasukavanje*, uz djelomično preklapanje s klasom 3 - *Udaranje u nepomični objekt*, što potvrđuje njihovu operativnu i okolišnu sličnost. U manjoj mjeri prisutno je i preklapanje između 5 - *Oštećenje / Gubitak opreme* i 7 - *Gubitak upravljivosti*.
Ukupno gledano, ansambl model pokazuje veću stabilnost i uravnoteženost u odnosu na pojedinačne modele, uz zadržavanje visoke točnosti kod ključnih događaja, čime predstavlja pouzdanu osnovu za daljnju primjenu slojevitih pristupa i završnu evaluaciju.

Tablica 55. Matrica zabune ansambl modela

Stvarna klasa → / Predikcija ↓	1	2	3	4	5	6	7
1	8	1	1	2	0	1	0
2	0	38	1	11	1	0	1
3	2	4	34	1	4	2	3
4	1	8	3	33	1	1	1
5	0	0	2	0	48	1	5
6	1	1	0	1	0	8	0
7	0	0	0	0	1	0	5

Izvor: izradio autor

Tablica 56 prikazuje metrike klasifikacije ansambl modela, koje potvrđuju njegovu uravnoteženu i stabilnu učinkovitost. Ostvarena točnost (0,73729), uz visoku specifičnost (0,95485) i negativnu prediktivnu vrijednost (0,95394), ukazuje na pouzdano razlikovanje

klasa, dok F1 mjera (0,68761) potvrđuje dobru ravnotežu između preciznosti i osjetljivosti. Iako ansambl model ne nadmašuje model slučajnih šuma u svim pokazateljima, ostvaruje stabilnije rezultate u odnosu na modele potpornih vektora i neuronskih mreža, čime se potvrđuje njegova robusnost i pouzdanost.

Tablica 56. Ključne klasifikacijske metrike ansambl modela

Model	Ansambl
Accuracy	0,73729
ErrorRate	0,26271
Sensitivity (makroprosjek)	0,73306
Specificity	0,95485
Precision (makroprosjek)	0,67652
NPV	0,95394
F1 (makroprosjek)	0,68761
FDR	0,32348
FOR	0,04606
FPR	0,04515
FNR	0,26694
LR_Positive	20,64744
LR_Negative	0,27957
DOR (makroprosjek)	82,27012

Izvor: izradio autor

Zaključno, ansambl model potvrđuje svoju vrijednost kroz stabilne i uravnotežene performanse, uz visoku sposobnost razlikovanja klasa i smanjen utjecaj pojedinačnih pogrešaka. Iako ne postiže najbolje rezultate u svim metrikama u odnosu na model slučajnih šuma, njegova robusnost i konzistentnost čine ga pouzdanom osnovom za daljnju nadogradnju modeliranja.

4.7. Slojeviti model

Na temelju prethodno utvrđenih prednosti i ograničenja pojedinačnih modela, u nastavku se primjenjuje slojeviti pristup (engl. *stacking*), koji omogućuje optimizirano kombiniranje predikcija baznih modela putem meta-modela te dodatno poboljšanje ukupnih performansi.

Tablica 57. prikazuje matricu zabune slojevitog modela. Model zadržava visoku točnost kod dominantnih klasa, posebno za 5 – *Oštećenje / Gubitak opreme* (48), dok su 2 – *Sudar* (38) i 4 – *Nasukavanje* (32) također vrlo dobro klasificirani. Klase s manjim brojem uzoraka, poput 1 – *Prevrtanje / Naginjanje* (6) i 7 – *Gubitak upravljivosti* (6), pokazuju nešto slabiju, ali i dalje stabilnu klasifikaciju. Najizraženije pogreške prisutne su između klasa 2 – *Sudar*, 4 – *Nasukavanje* i 3 – *Udaranje u nepomični objekt*, što potvrđuje njihovu međusobnu sličnost, dok se manja preklapanja javljaju između 5 – *Oštećenje / Gubitak opreme* i 7 – *Gubitak upravljivosti*.

Tablica 57. Matrica zabune slojevitog modela

Stvarna klasa → / Predikcija ↓	1	2	3	4	5	6	7
1	6	0	1	0	0	0	0
2	1	38	4	13	0	0	1

3	1	3	34	1	3	3	3
4	1	10	1	32	1	1	1
5	0	0	1	0	48	0	4
6	3	1	0	1	0	9	0
7	0	0	0	1	3	0	6

Izvor: izradio autor

Metrike prikazane u Tablici 58. potvrđuju uravnotežene performanse slojevitog modela. Ostvarena točnost (0,73305), visoka specifičnost (0,95394) i NPV (0,95322) ukazuju na pouzdano razlikovanje klasa, dok F1 mjera (0,68602) potvrđuje dobru ravnotežu između preciznosti i osjetljivosti. Iako su rezultati vrlo slični ansambl modelu, slojeviti model ne donosi značajno poboljšanje u ukupnoj točnosti, ali pokazuje blago povećanje diskriminacijske snage (DOR).

Tablica 58. Ključne klasifikacijske metrike slojevitog modela

Model	Slojeviti
Accuracy	0,73305
ErrorRate	0,26695
Sensitivity	0,72307
Specificity	0,95394
Precision	0,67025
NPV	0,95322
F1	0,68602
FDR	0,32975
FOR	0,04678
FPR	0,04606
FNR	0,27693
LR_Positive	20,38946
LR_Negative	0,29119
DOR (makroprosjek)	101,15316

Izvor: izradio autor

Slojeviti model potvrđuje stabilne i pouzdane performanse, ali bez značajno boljih rezultata u odnosu na ansambl pristup. Njegova glavna vrijednost ogleda se u dodatnoj optimizaciji kombiniranja pojedinačnih modela, što ga čini pogodnim za završnu usporednu analizu svih primijenjenih pristupa, koja obuhvaća usporedbu matrica procijenjenih vjerojatnosti pojave glavnih događaja (Prilog B, Tablice B5. i B6.) te usporedbu matrica rizika dobivenih kombiniranjem tih vjerojatnosti s pripadajućim posljedicama.

4.8. Integrirana usporedna analiza klasifikacijskih performansi, vjerojatnosti i procjene rizika modela

U ovom poglavlju prikazan je objedinjeni pregled rezultata svih primijenjenih modela radi cjelovite evaluacije njihove klasifikacijske uspješnosti, stabilnosti i primjenjivosti u predikciji glavnih događaja pomorskih nezgoda te procjeni rizika na temelju procijenjenih vjerojatnosti njihove pojave.

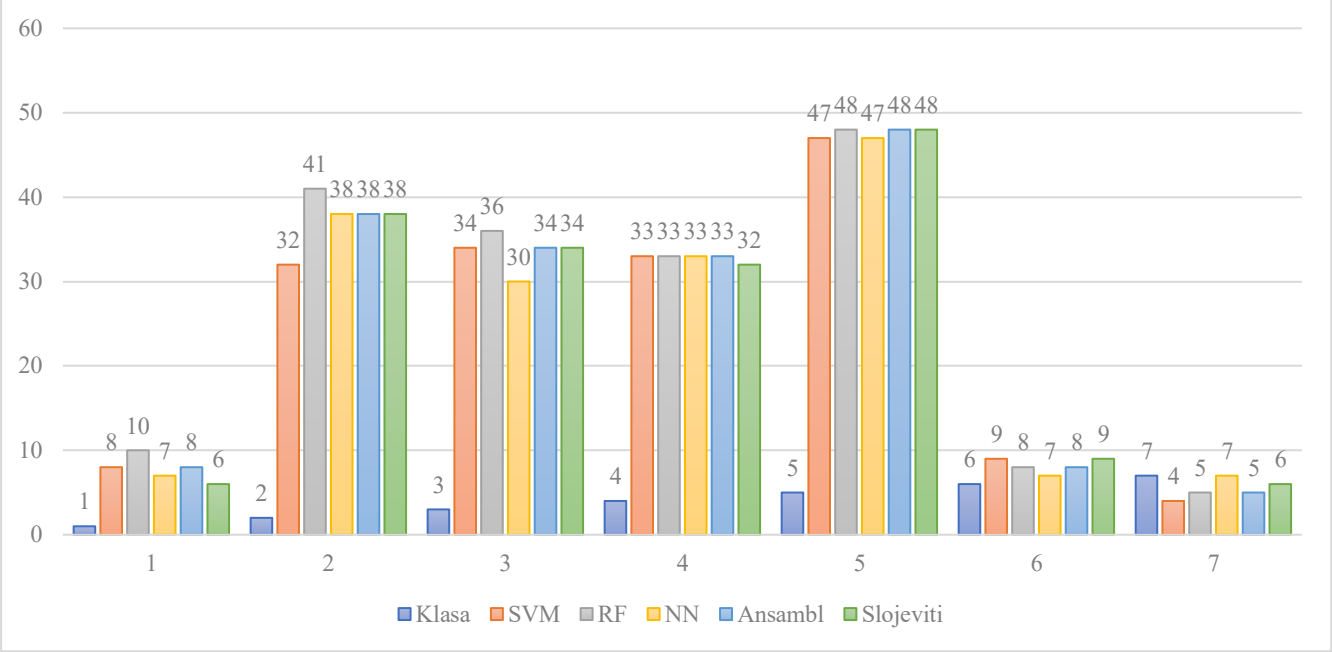
Tablica 59. i Slika 24. prikazuju usporedbu broja točno klasificiranih slučajeva prema glavnim događajima (klasa) i modelima. Model slučajnih šuma ostvaruje najveći ukupan broj

točnih klasifikacija (181), pri čemu se posebno ističe u klasama 2 – *Sudar* i 1 – *Prevrtanje / Naginjanje*. Ansambl model (174) i slojeviti model (173) postižu vrlo slične rezultate, uz stabilne performanse kroz većinu klasa. Kod klasa 2 – *Sudar*, 3 – *Udaranje u nepomični objekt*, 4 – *Nasukavanje* i 5 – *Oštećenje / Gubitak* opreme svi modeli ostvaruju velik broj točnih klasifikacija, dok su razlike izraženije kod slabije zastupljenih klasa, osobito kod klase 7 – *Gubitak upravljivosti*. Vizualni prikaz na Slici 25. dodatno potvrđuje ove odnose i jasno ističe blagu prednost modela slučajnih šuma te ujednačenost rezultata ansambl i slojevitog modela.

Tablica 59. Usporedni prikaz pravilno klasificiranih slučajeva po klasama i modelima

Klasa	SVM	RF	NN	Ansambl	Slojeviti
1	8	10	7	8	6
2	32	41	38	38	38
3	34	36	30	34	34
4	33	33	33	33	32
5	47	48	47	48	48
6	9	8	7	8	9
7	4	5	7	5	6
Ukupno	167	181	169	174	173

Izvor: izradio autor



Slika 24. Usporedni prikaz točno klasificiranih slučajeva po svim klasama i svim modelima (SVM, RF, NN, ansambl i slojeviti)

Izvor: izradio autor

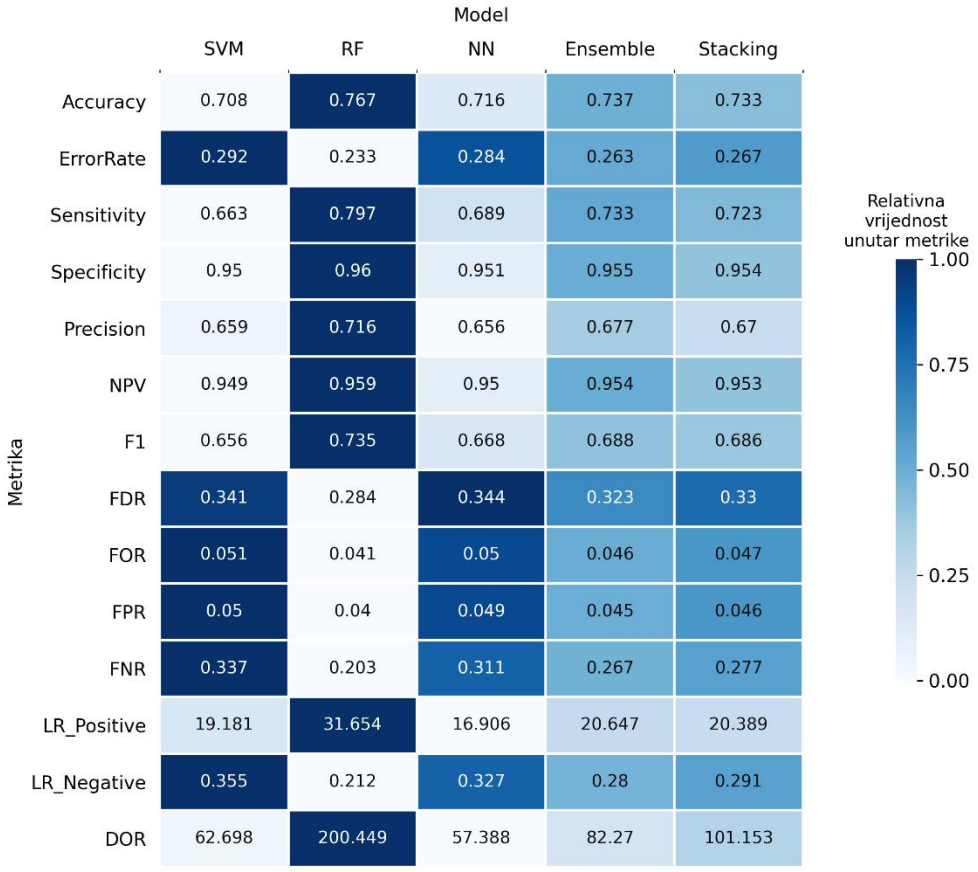
Usporedba metrika u Tablici 60. potvrđuje nadmoć modela slučajnih šuma u većini pokazatelja, osobito u točnosti (0,76695), osjetljivosti i omjeru izgleda (200,45), što ukazuje na njegovu visoku diskriminacijsku sposobnost. Ansambl model postiže stabilne i uravnotežene vrijednosti kroz sve metrike, dok slojeviti model pokazuje gotovo identične performanse, uz blago povećanje diskriminacijske snage (101,15) u odnosu na ansambl. Model potpornih vektora i neuronskih mreža ostvaruju niže vrijednosti većine metrika, uz izraženiju varijabilnost rezultata.

Tablica 60. Usporedni prikaz metrika uspješnosti klasifikacijskih modela

Model	SVM	RF	NN	Ansambl	Slojeviti
Accuracy	0,70763	0,76695	0,71610	0,73729	0,73305
ErrorRate	0,29237	0,23305	0,28390	0,26271	0,26695
Sensitivity (makroprosjek)	0,66335	0,79694	0,68854	0,73306	0,72307
Specificity	0,94971	0,96000	0,95094	0,95485	0,95394
Precision (makroprosjek)	0,65891	0,71554	0,65614	0,67652	0,67025
NPV	0,94915	0,95882	0,95040	0,95394	0,95322
F1 (makroprosjek)	0,65636	0,73488	0,66822	0,68761	0,68602
FDR	0,34109	0,28446	0,34386	0,32348	0,32975
FOR	0,05085	0,04118	0,04960	0,04606	0,04678
FPR	0,05029	0,04000	0,04906	0,04515	0,04606
FNR	0,33665	0,20306	0,31146	0,26694	0,27693
LR_Positive	19,18106	31,65425	16,90558	20,64744	20,38946
LR_Negative	0,35454	0,21246	0,32733	0,27957	0,29119
DOR (makroprosjek)	62,69772	200,44876	57,38777	82,27012	101,15316

Izvor: izradio autor

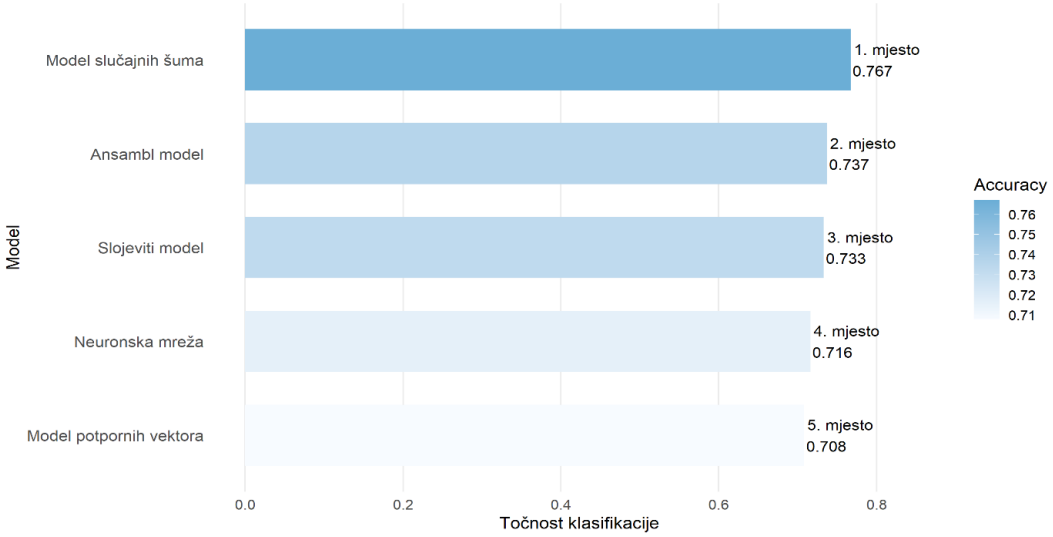
Slika 25. dodatno potvrđuje navedene odnose, pri čemu je vidljivo da model slučajnih šuma dominira u ključnim metrikama klasifikacije glavnih događaja, dok ansambl i slojeviti modeli pružaju kompromis između stabilnosti i točnosti. Njihova vrijednost posebno dolazi do izražaja u kontekstu daljnje analize procijenjenih vjerojatnosti pojave glavnih događaja (Prilog B, Tablice B5. i B6.) i na temelju njih izvedenih matrica rizika, gdje omogućuju konzistentnije i robusnije rezultate u odnosu na pojedinačne modele. Intenzitet boje prikazuje relativni položaj vrijednosti modela unutar pojedine metrike. Svjetlija nijansa označava manju, a tamnija veću vrijednost u promatranom retku. Brojevi u poljima predstavljaju izvorne vrijednosti metrika. Tamnija nijansa ne označava nužno bolji rezultat jer je kod pojedinih metrika manja vrijednost poželjnija.



Slika 25. Usporedni prikaz metrika uspješnosti klasifikacijskih modela

Izvor: izradio autor

Slika 26. prikazuje usporedni poredak modela prema ukupnoj točnosti klasifikacije. Najbolji rezultat ostvaruje model slučajnih šuma, nakon čega slijede ansambl model i slojeviti model. Neuronska mreža i model potpornih vektora ostvaruju nešto nižu točnost, što potvrđuje prednost modela slučajnih šuma u ukupnoj klasifikacijskoj uspješnosti.



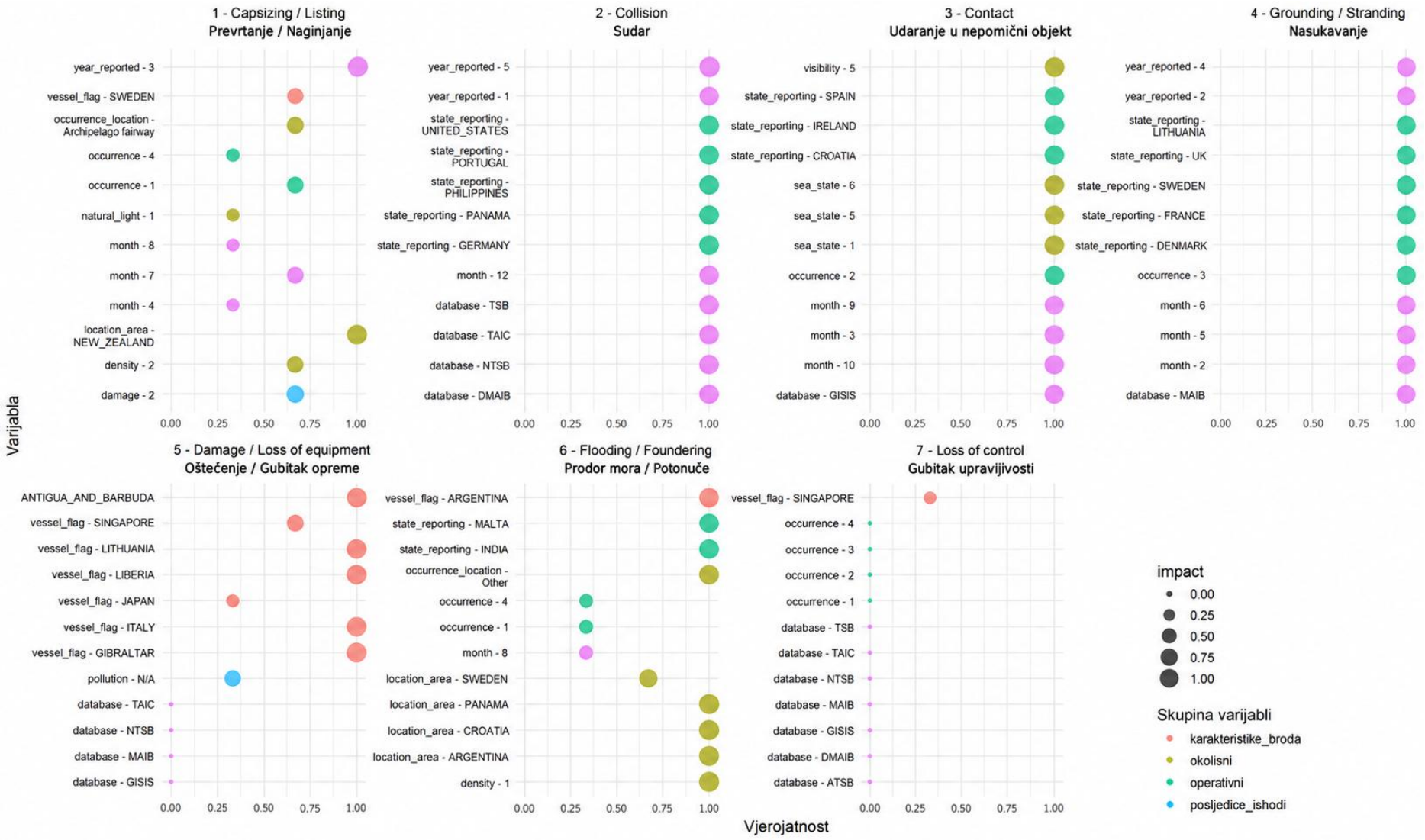
Slika 26. Poredak modela prema ukupnoj točnosti

Izvor: izradio autor

Iako slojeviti model ne nadmašuje ansambl model niti model slučajnih šuma u ukupnoj točnosti, potvrđuje stabilnost i pouzdanost u kombiniranju predikcija, čime zajedno s ansambl modelom predstavlja važnu osnovu za predstavlja važnu osnovu za integriranu analizu procijenjenih vjerojatnosti pojave glavnih događaja i njihovu primjenu u okviru procjene rizika nakon uključivanja indeksa posljedica.
Nakon provedene analize utjecaja varijabli na procijenjene vjerojatnosti pojave glavnih događaja pomorskih nezgoda za pojedinačne modele (model potpornih vektora, model slučajnih šuma i neuronskih mreža), u nastavku je prikazana usporedna interpretacija rezultata za ansambl i slojeviti model. Cilj je utvrditi u kojoj mjeri ovi složeniji pristupi doprinose stabilnijoj procjeni vjerojatnosti pojave glavnih događaja te jasnijem prepoznavanju ključnih skupina varijabli.

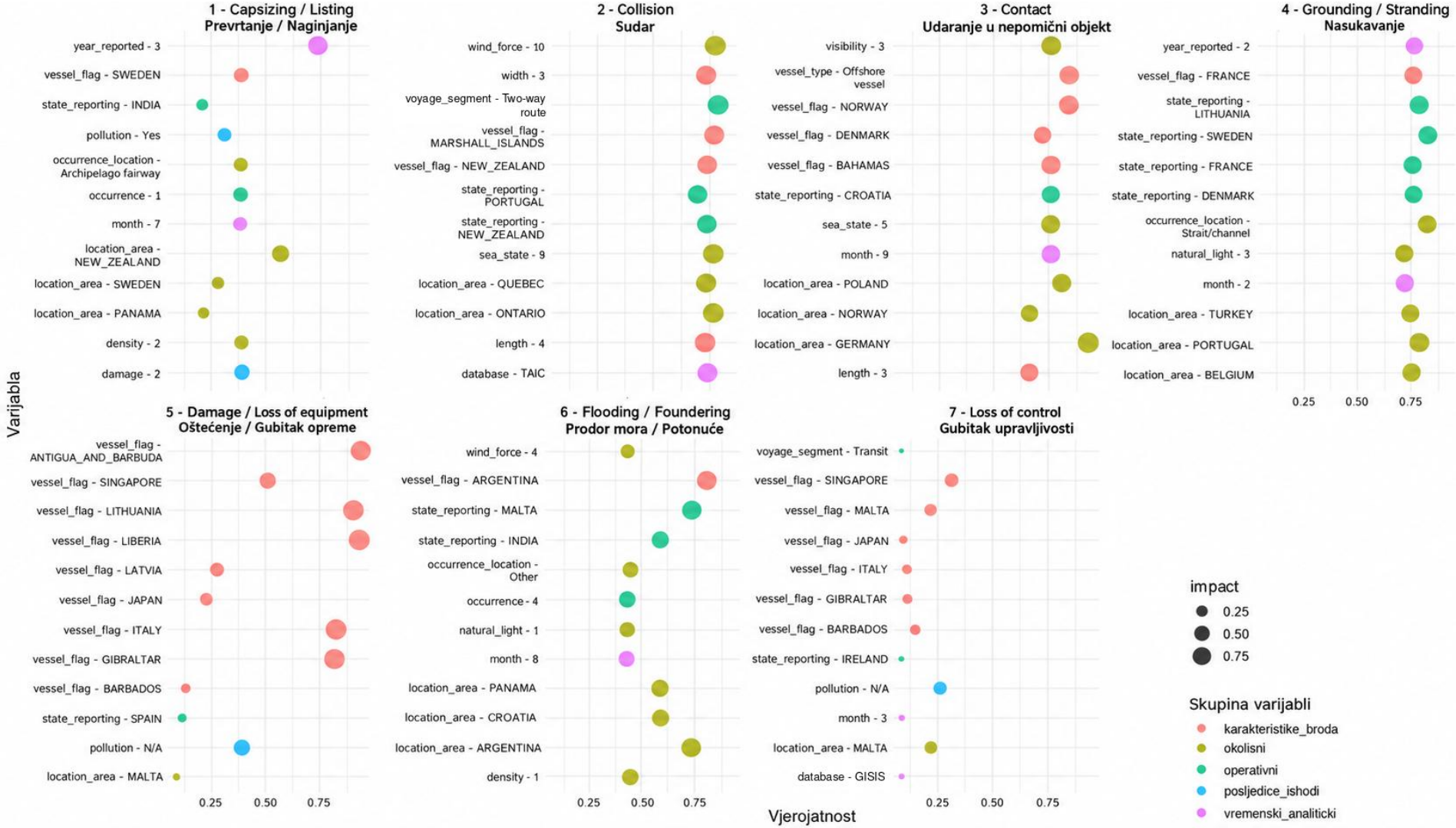
Kao što je prikazano na Slici 27., ansambl model pokazuje izraženu koncentraciju vjerojatnosti unutar dominantnih klasa *Sudar*, *Udaranje u nepomični objekt* i *Nasukavanje*. Najveći utjecaj dolazi iz *okolišnih čimbenika* (*stanje mora*, *vidljivost*, *jačina vjetra*, *vrijeme, gustoća prometa, područje nezgode*) te *operativnih čimbenika (segment plovidbe i država izvjestiteljica) koji* dodatno diferenciraju klasu *Sudar*. Procijenjene vjerojatnosti pojave glavnih događaja u pravilu su visoko koncentrirane unutar pojedine klase. Kod rjeđih klasa (*Prevrtanje / Naginjanje*, *Prodor mora / Potonuće* i *Gubitak upravljivosti*) utjecaji su slabije izraženi i raspodijeljeni kroz više skupina varijabli.
S druge strane, Slika 28. pokazala je da je slojeviti model zadržao iste dominantne skupine varijabli, uz znatno izraženiju raspodjelu procijenjenih vjerojatnosti pojave glavnih događaja između više klasa. Uz *uvjete plovidbe* i *lokacijska obilježja*, veći značaj dobile su i tehničke karakteristike broda *(tip, kategorija, zastava, dimenzije, materijal, starost),* što je ukazivalo na dublju razinu modeliranja odnosa u podacima. Međutim, takva raspodjela rezultirala je i većim preklapanjem između glavnih događaja, osobito kod *Sudara, Udaranja u nepomični objekt i Nasukavanja*.

Usporedno gledano, ansambl model osigurao je stabilniju i robusniju procjenu vjerojatnosti glavnog događaja s jasnijom klasnom diferencijacijom, dok je slojeviti model pružio detaljniji uvid u međusobne odnose varijabli i preklapanja među klasama. Time se ansambl model pokazao pogodnijim za konzistentnu procjenu vjerojatnosti pojave glavnih događaja i njihovu primjenu u okviru procjene rizika, dok je slojeviti model imao veću interpretativnu vrijednost u analizi složenih obrazaca ponašanja sustava.



Slika 27. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda primjenom ansambl modela

Izvor: izradio autor



Slika 28. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda primjenom slojevitog modela

Izvor: izradio autor

U nastavku je usporedni prikaz matrica rizika dobivenih kombiniranjem procijenjenih vjerojatnosti i težina posljedica, s ciljem uvida u razlike između ansambl i slojevitog modela u procjeni razine rizika.

Kao što je prikazano na Slici 29., matrica rizika temeljena na ansambl modelu pokazuje jasnu i stabilnu raspodjelu rizika, pri čemu se dominantni događaji (*Sudar*, *Nasukavanje* i *Udaranje u nepomični objekt*) uglavnom svrstavaju u srednje do velike razine rizika. Rjeđi događaji (*Prevrtanje / Naginjanje, Prodor mora / Potonuće i Gubitak upravljivosti*) zadržavaju male ili srednje razine rizika, pri čemu njihova konačna pozicija proizlazi iz kombinacije procijenjene vjerojatnosti pojave i razine posljedica. Posebno se kod događaja s potencijalno ozbiljnim posljedicama, poput *Prodora mora / Potonuća*, razina rizika ne određuje samo učestalošću pojave, nego i težinom mogućih posljedica.

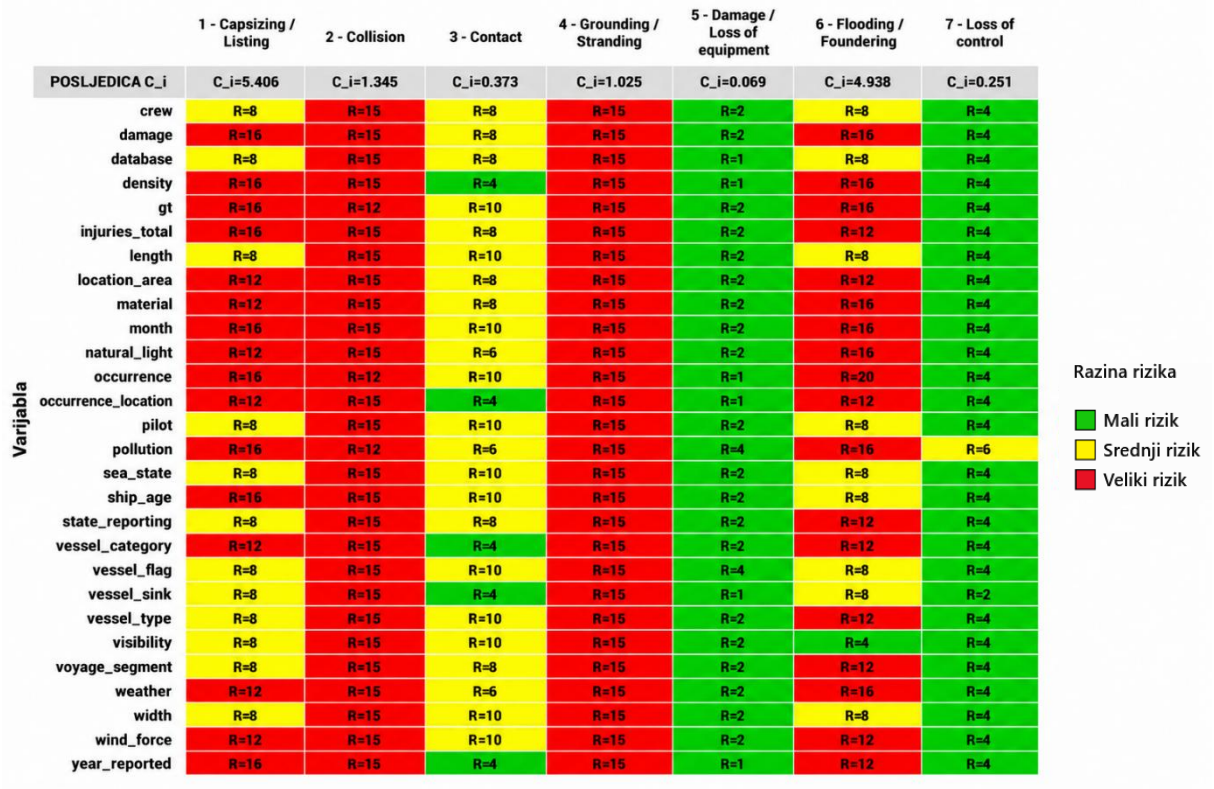
Varijabla	1 - Capsizing / Listing	2 - Collision	3 - Contact	4 - Grounding / Stranding	5 - Damage / Loss of equipment	6 - Flooding / Foundering	7 - Loss of control
POSLJEDICA C j	C j=5.406	C j=1.345	C j=0.373	C j=1.025	C j=0.069	C j=4.938	C j=0.251
crew	R=4	R=12	R=10	R=15	R=1	R=4	R=2
damage	R=16	R=3	R=10	R=15	R=1	R=12	R=2
database	R=4	R=15	R=8	R=15	R=1	R=4	R=2
density	R=16	R=3	R=2	R=15	R=1	R=20	R=2
gt	R=12	R=3	R=10	R=15	R=1	R=12	R=2
injuries_total	R=4	R=3	R=10	R=15	R=1	R=4	R=2
length	R=4	R=15	R=10	R=15	R=1	R=4	R=2
location_area	R=8	R=15	R=10	R=15	R=1	R=16	R=2
material	R=12	R=12	R=10	R=15	R=1	R=12	R=2
month	R=16	R=12	R=10	R=15	R=1	R=12	R=2
natural_light	R=16	R=3	R=8	R=15	R=1	R=16	R=2
occurrence	R=20	R=3	R=10	R=15	R=1	R=16	R=2
occurrence_location	R=16	R=3	R=2	R=15	R=1	R=16	R=2
pilot	R=4	R=3	R=10	R=15	R=1	R=4	R=2
pollution	R=16	R=3	R=10	R=15	R=4	R=16	R=2
sea_state	R=4	R=15	R=10	R=15	R=1	R=4	R=2
ship_age	R=4	R=15	R=10	R=15	R=1	R=4	R=2
state_reporting	R=4	R=15	R=8	R=15	R=1	R=12	R=2
vessel_category	R=12	R=15	R=6	R=15	R=1	R=12	R=2
vessel_flag	R=8	R=15	R=10	R=15	R=4	R=8	R=4
vessel_sink	R=4	R=12	R=2	R=15	R=1	R=4	R=2
vessel_type	R=8	R=15	R=10	R=15	R=1	R=8	R=2
visibility	R=4	R=3	R=10	R=15	R=1	R=4	R=2
voyage_segment	R=8	R=15	R=8	R=15	R=1	R=8	R=2
weather	R=12	R=12	R=6	R=15	R=1	R=12	R=2
width	R=4	R=15	R=10	R=15	R=1	R=4	R=2
wind_force	R=12	R=15	R=10	R=15	R=1	R=8	R=2
year_reported	R=20	R=15	R=2	R=15	R=1	R=4	R=2

Razina rizika
Mali rizik
Srednji rizik
Veliki rizik

Slika 29. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih primjenom ansambl modela i pripadajućih indeksa posljedica

Izvor: izradio autor

Slika 30. pokazuje da slojeviti model daje nešto disperzniju raspodjelu rizika. Iako su dominantne klase i dalje prepoznate kao ključni izvori rizika, uočava se veće preklapanje između razina rizika, osobito kod operativno sličnih događaja. Takva raspodjela proizlazi iz razlika u procijenjenim vjerojatnostima pojave glavnih događaja između modela, dok su razine posljedica određene prema prethodno definiranoj metodologiji procjene rizika. Usporedno gledano, ansambl model omogućuje konzistentniju i pregledniju klasifikaciju razina rizika, dok slojeviti model pruža detaljniji, ali manje homogen prikaz. Time se ansambl model pokazuje pogodnijim za operativnu primjenu i donošenje odluka zbog stabilnije raspodjele procijenjenih razina rizika, dok slojeviti model pruža veću analitičku vrijednost u razumijevanju varijabilnosti procijenjenih vjerojatnosti i njihovog utjecaja na strukturu rizika.



Slika 30. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih primjenom slojevitog modela i pripadajućih indeksa posljedica

Izvor: izradio autor

Tablica 61. prikazuje usporedne rezultate vjerojatnosti pojave glavnih događaja i pripadajućih razina rizika dobivenih primjenom pet modela strojnog učenja na istom skupu podataka. Uočava se da neuronske mreže i slojeviti model generiraju najveći udio velikih rizika, što ukazuje na veću sklonost tih modela dodjeljivanju velikih razina rizika sigurnosno kritičnim situacijama. Nasuprot tome, primjenom ansambl modela dobiva se najveći udio malih rizika i najniža prosječna vrijednost indeksa rizika, što upućuje na stabilniju i umjereniju procjenu.

Vrijednosti ukupnog doprinosa riziku (ΣR_i) potvrđuju razlike među modelima, pri čemu su najviše procjene dobivene primjenom neuronskih mreža i modela potpornih vektora, dok ansambl model daje najnižu ukupnu procjenu rizika. Model slučajnih šuma i slojeviti model zauzimaju srednji položaj, što ukazuje na različitu raspodjelu procijenjenih vjerojatnosti i njihov utjecaj na izvedenu razinu rizika.

Razlike među modelima dodatno se očituju u maksimalnim vrijednostima procijenjenih vjerojatnosti pojave glavnih događaja, koje ukazuju na različite obrasce koncentracije predikcija pojedinih modela. U cjelini, rezultati potvrđuju da izbor modela utječe na izvedene razine rizika, pri čemu pojedini modeli naglašavaju sigurnosno kritične scenarije, dok drugi daju stabilnije i umjerenije raspodjele procijenjenih vrijednosti.

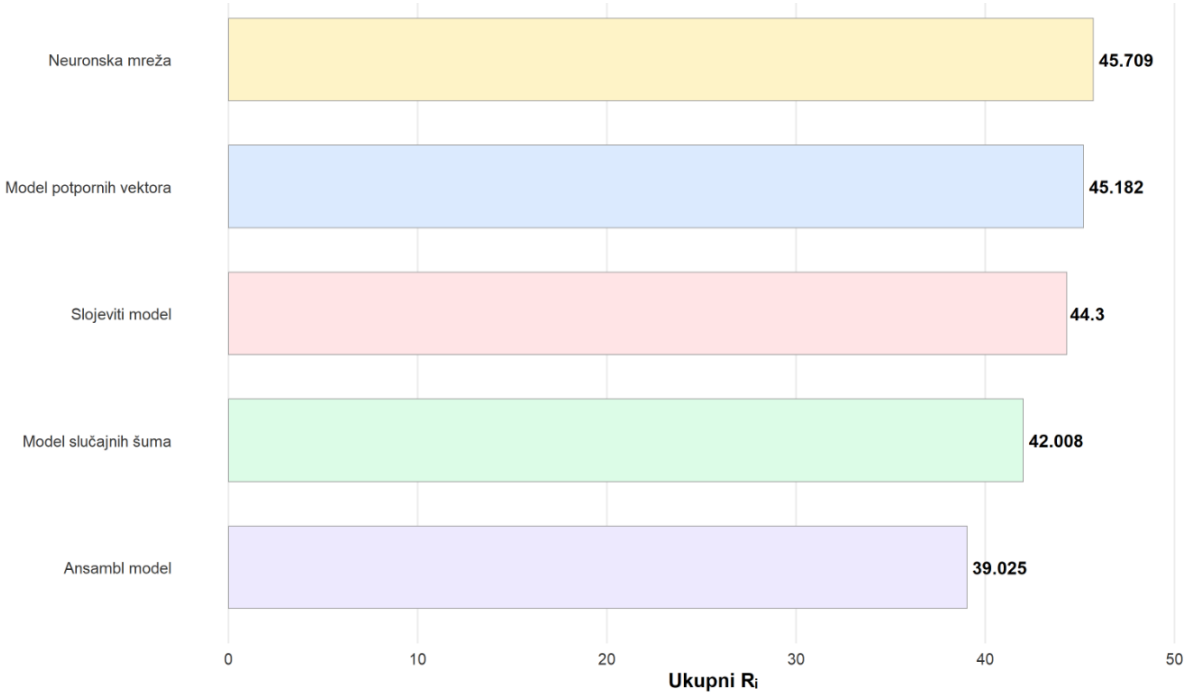
Tablica 61. Usporedni prikaz procijenjenih vjerojatnosti pojave glavnih događaja i izvedenih razina rizika za modele strojnog učenja

Model	SVM	RF	NN	Ansambl	Slojeviti
Broj_malih_rizika	64	55	68	92	61

Broj_srednjih_rizika	46	53	36	32	45
Broj_velikih_rizika	86	88	92	72	90
Prosjecni_risk_score	9,230	9,378	9,276	7,964	9,571
Medijan_risk_score	10	8	10	8	10
Maksimalni_risk_score	20	16	20	20	20
Ukupni_R_i	45,182	42,008	45,709	39,025	44,300
Prosjecna_P	0,143	0,143	0,143	0,143	0,143
Maksimalna_P	0,593	0,523	0,667	0,833	0,561
Ukupno	196	196	196	196	196
Udio_malih_rizika	0,327	0,281	0,347	0,469	0,311
Udio_srednjih_rizika	0,235	0,270	0,184	0,163	0,230
Udio_velikih_rizika	0,439	0,449	0,469	0,367	0,459

Izvor: izradio autor

Slika 31. prikazuje ukupni doprinos riziku pomorskih nezgoda izražen kao suma umnožaka vjerojatnosti pojave glavnih događaja i težine posljedica (ΣR_i), dobiven primjenom različitih modela strojnog učenja. Uočeno je da su neuronske mreže i model potpornih vektora rezultirali najvišim vrijednostima ukupnog doprinosa riziku, dok je najniža vrijednost ostvarena primjenom ansambl modela. Model slučajnih šuma i slojeviti model zauzeli su srednji položaj. Dobivene razlike ukazuju na to da različiti modeli, zbog različitih raspodjela procijenjenih vjerojatnosti glavnih događaja, dovode do različitih ukupnih vrijednosti rizika nakon uključivanja težine posljedica. Najviši zbroj doprinosa riziku iznosi 45,709 za model neuronske mreže, a najniži 39,025 za ansambl model. Apsolutna razlika iznosi 6,684, odnosno približno 17,1 % u odnosu na najnižu vrijednost. Razlika je deskriptivna, njezina statistička značajnost i operativno značenje nisu zasebno utvrđeni. Manji zbroj sam po sebi ne predstavlja dokaz veće točnosti modela ni povoljnijeg stvarnog sigurnosnog ishoda.

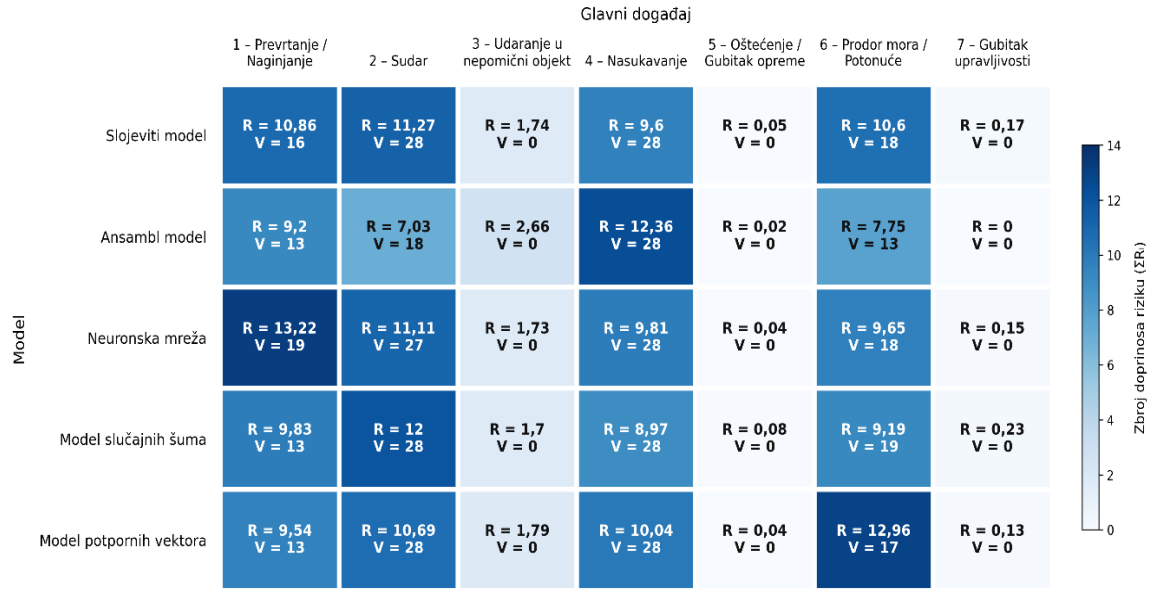


Slika 31. Usporedba ukupnog doprinosa riziku (ΣR_i) po modelima strojnog učenja

Izvor: izradio autor

Slika 32. prikazuje raspodjelu agregiranih doprinosa riziku prema sedam glavnih događaja za sve promatrane modele. Više vrijednosti uočavaju se kod klasa 1 – *Prevrtanje / Naginjanje*, 2 – *Sudar*, 4 – *Nasukavanje* i 6 – *Prodor mora / Potonuće*, pri čemu se njihov iznos razlikuje među modelima. Najveća prikazana vrijednost ostvarena je kod klase 1 – *Prevrtanje / Naginjanje* primjenom modela neuronske mreže (R = 13,22).

Parcijalni doprinos riziku izračunava se kao $R_i = P_i \cdot C_i$, gdje je P_i procijenjena vjerojatnost glavnog događaja, a C_i pripadajući indeks posljedica. Vrijednost R u pojedinom polju predstavlja zbroj doprinosa riziku (ΣR_i) za pripadajući model i glavni događaj. Oznaka V prikazuje broj kombinacija varijable i glavnog događaja razvrstanih u kategoriju velikog rizika. Tamnija nijansa označava veću vrijednost zbroja doprinosa riziku.

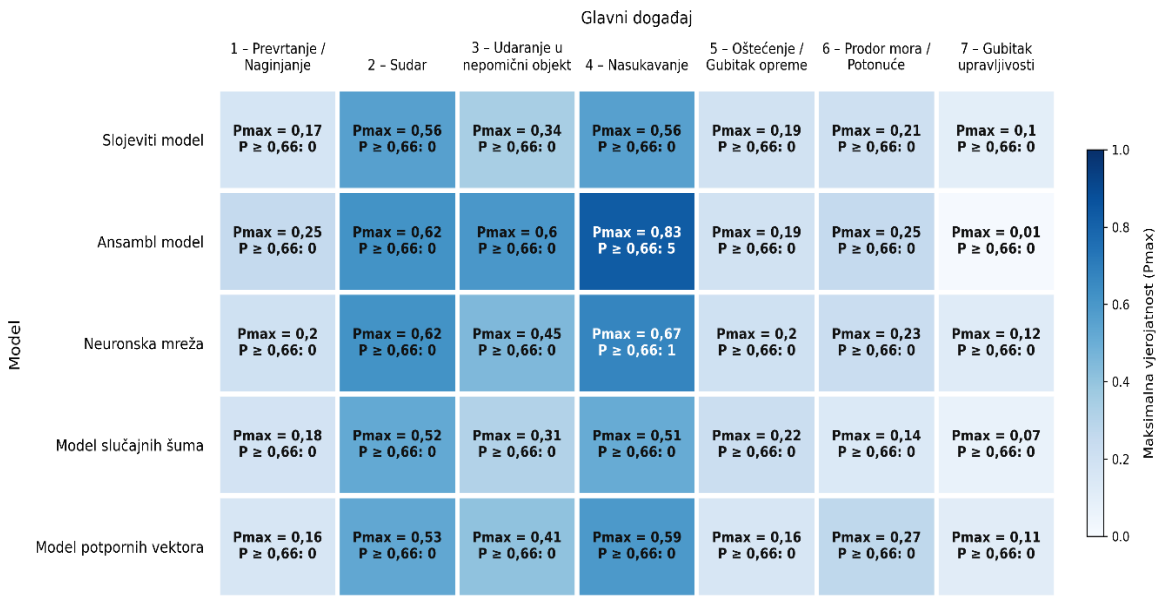


Slika 32. Toplinska matrica raspodjele rizika po glavnim događajima i modelima

Izvor: izradio autor

Slika 33. prikazuje najveće prosječne procijenjene vjerojatnosti pripadnosti pojedinim klasama glavnih događaja pomorskih nezgoda za svaki promatrani model. Prosječne vjerojatnosti najprije su izračunate za pojedine varijable, a zatim je za svaki model i klasu izdvojena najveća vrijednost (*Pmax*). Više vrijednosti uočavaju se kod klasa 2 – *Sudar*, 3 – *Udaranje u nepomični objekt* i 4 – *Nasukavanje*. Najveća prikazana vrijednost ostvarena je kod ansambl modela za klasu 4 – *Nasukavanje* (*Pmax = 0,83*).

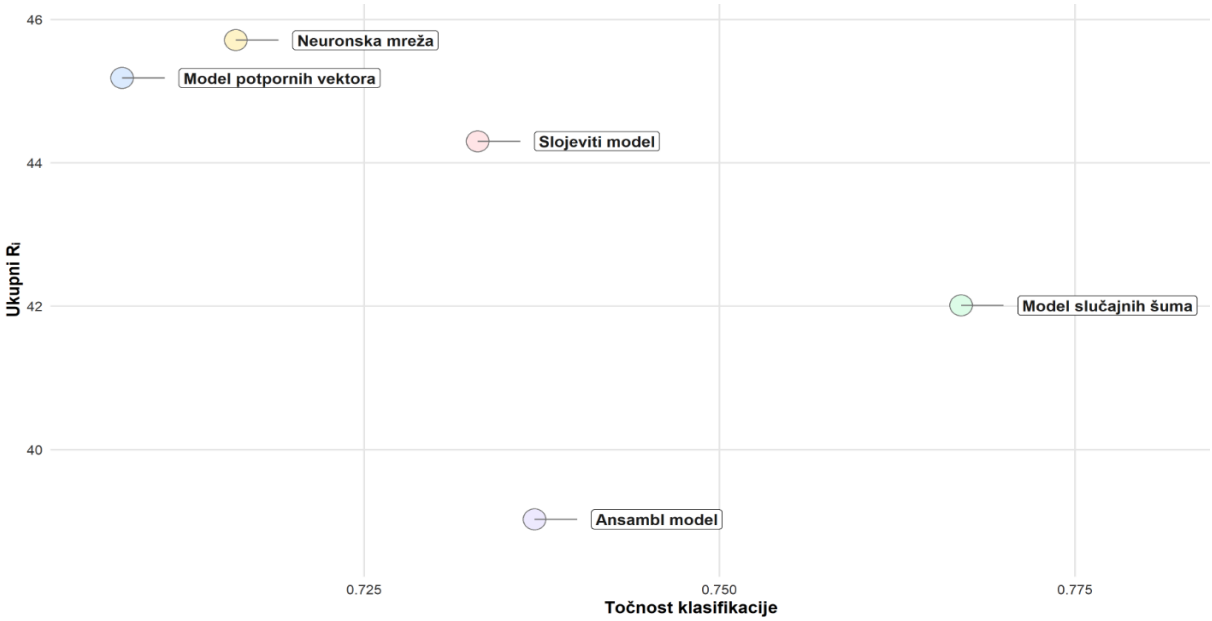
U svakom polju prikazan je i broj varijabli čija je prosječna procijenjena vjerojatnost za promatranu klasu jednaka ili veća od praga $P \geq 0{,}66$. Za klasu 4 – *Nasukavanje* taj broj iznosi 5 kod ansambl modela i 1 kod modela neuronske mreže, dok je u ostalim prikazanim poljima jednak 0. Prelazak odabranog praga označava visoku procijenjenu vjerojatnost, ali sam po sebi ne potvrđuje točnost klasifikacije. Tamnija nijansa označava veću vrijednost *Pmax*.



Slika 33. Toplinska matrica raspodjele vjerojatnosti po glavnim događajima i modelima

Izvor: izradio autor

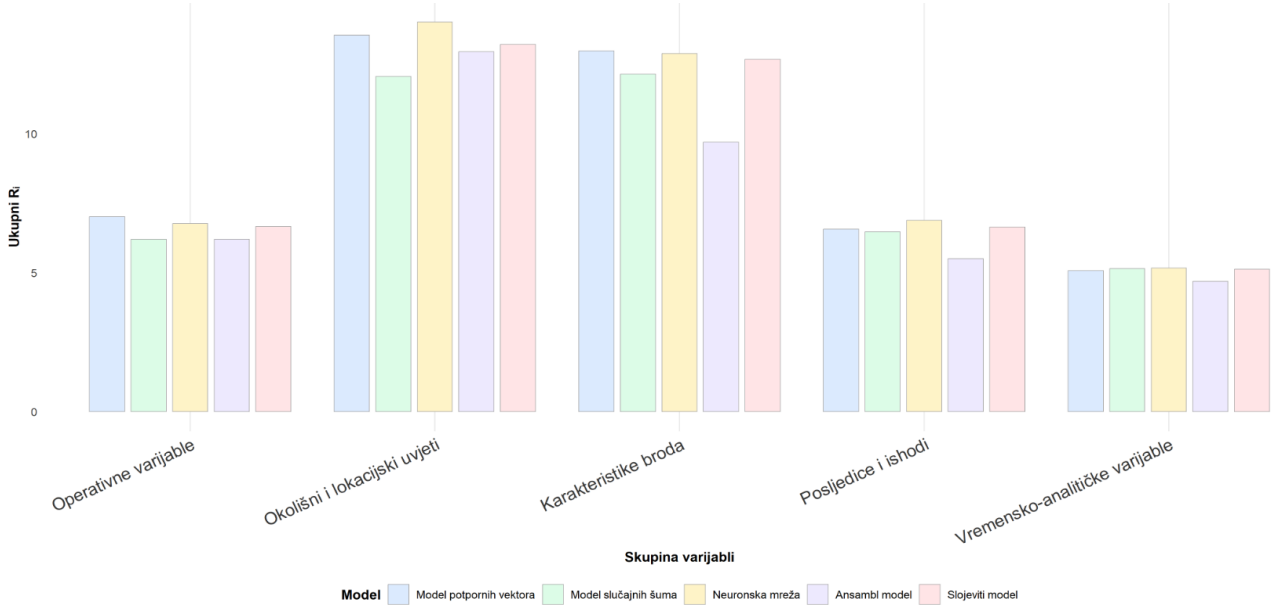
Slika 34. prikazuje odnos između klasifikacijske točnosti modela i ukupnog procijenjenog rizika pomorskih nezgoda. Uočava se da viša klasifikacijska točnost ne podrazumijeva nužno i višu procijenjenu razinu rizika, što je posebno vidljivo kod modela slučajnih šuma, koji ostvaruje visoku točnost uz umjereniju razinu procijenjenog rizika. Nasuprot tome, neuronske mreže i slojeviti model rezultiraju višim procijenjenim razinama rizika uz nešto nižu klasifikacijsu točnost, što ukazuje na izraženiju osjetljivost u prepoznavanju sigurnosno kritičnih situacija. Ovakvi rezultati ukazuju na postojanje kompromisa između prediktivne točnosti modela i ukupno procijenjene razine rizika, pri čemu na konačnu procjenu rizika, osim klasifikacijskih sposobnosti modela, utječu i pripadajući indeksi posljedica.



Slika 34. Odnos klasifikacijske točnosti (ACC) i ukupnog doprinosa riziku (ΣR_i)

Izvor: izradio autor

Slika 35. prikazuje raspodjelu ukupnog procijenjenog rizika pomorskih nezgoda prema skupinama varijabli. Uočava se da najveći doprinos riziku proizlazi iz okolišnih i lokacijskih čimbenika te karakteristika broda, što potvrđuje njihovu ključnu ulogu u nastanku nezgoda. Operativni i vremensko-analitički čimbenici imaju manji, ali i dalje značajan utjecaj. Razlike među modelima očituju se u intenzitetu procjene unutar pojedinih skupina, pri čemu neuronska mreža i slojeviti model rezultiraju većim procijenjenim vrijednostima, dok ansambl model daje ujednačeniju raspodjelu.



Slika 35. Usporedba doprinosa riziku (ΣR_i) prema skupinama varijabli i modelima

Izvor: izradio autor

Zaključno, provedena analiza pokazuje da primjena različitih modela strojnog učenja rezultira različitim procijenjenim razinama rizika pomorskih nezgoda, pri čemu se jasno izdvajaju modeli s većom osjetljivošću na sigurnosno kritične situacije i modeli koji daju stabilnije, umjerenije procjene. Dobiveni rezultati potvrđuju da ukupni rizik proizlazi iz međudjelovanja vjerojatnosti i težine posljedica te da izbor modela ima značajan utjecaj na konačnu procjenu sigurnosnog profila obalne plovidbe.

U cilju dodatne provjere pouzdanosti i primjenjivosti dobivenih rezultata, u nastavku rada provodi se validacija modela, kojom se ocjenjuje njihova prediktivna uspješnost i stabilnost na neviđenim podacima.

4.9. Provjera ponašanja modela pri kontroliranoj promjeni ciljnih oznaka

Nakon treniranja, optimizacije hiperparametara i evaluacije razvijenih modela strojnog učenja provedena je dodatna analiza učinka promjene referentnih oznaka na rezultate klasifikacije. Takav pristup odabran je jer se u izvješćima i bazama podataka o pomorskim nezgodama mogu pojaviti pogreške pri klasifikaciji događaja, nekonzistentnosti i razlike u kriterijima izvještavanja.

Pogreške u oznakama klasa mogu utjecati na rezultate klasifikacije, pri čemu njihov učinak može ovisiti o načinu nastanka i udjelu pogrešnih oznaka (Frénay & Verleysen, 2014; Jiang et al., 2020). Za dodatnu provjeru odabrano je 19 zapisa iz testnog skupa od 236 zapisa,

što čini 8,05 % testnog skupa. Model potpornih vektora, model slučajnih šuma, neuronska mreža i ansambl model ispravno su klasificirali svih 19 odabranih zapisa, dok je slojeviti model ispravno klasificirao 18 zapisa. Na odabrane zapise primijenjena je unaprijed definirana shema promjene oznaka ciljne varijable *glavni događaj*, prikazana u Tablici 62. Prema toj shemi oznake klasa 1 i 2, 3 i 4 te 5 i 6 međusobno su zamijenjene, dok oznaka klase 7 nije mijenjana. Ulazne varijable pritom su ostale nepromijenjene. Od 19 odabranih zapisa, oznake su promijenjene kod 17 zapisa, dok su dva zapisa klase 7 zadržala izvornu oznaku. Promjene nisu bile ravnomjerno raspoređene među klasama. Iz klase 1 u klasu 2 prebačena su tri zapisa, iz klase 2 u klasu 1 četiri zapisa, između klasa 3 i 4 po dva zapisa u svakom smjeru, a između klasa 5 i 6 po tri zapisa u svakom smjeru.

Odabrana shema predstavlja jedan kontrolirani eksperimentalni scenarij, pri čemu su parovi klasa odabrani radi jednostavne i pregledne provedbe. Nije utvrđeno da su odabrani parovi klasa međusobno sličniji ili prikladniji od drugih mogućih zamjena.

Tablica 62. Shema kontrolirane izmjene ciljnih vrijednosti varijable *glavni događaj*

Prije izmjene	Nakon izmjene
1 – Prevrtanje / Naginjanje	2 – Sudar
2 – Sudar	1 – Prevrtanje / Naginjanje
3 – Udaranje u nepomični objekt	4 – Nasukavanje
4 – Nasukavanje	3 – Udaranje u nepomični objekt
5 – Oštećenje / Gubitak opreme	6 – Prodor mora / Potonuće
6 – Prodor mora / Potonuće	5 – Oštećenje / Gubitak opreme
7 – Gubitak upravljivosti	7 – Gubitak upravljivosti

Izvor: izradio autor

Na taj način stvoren je kontrolirani eksperimentalni scenarij u kojem modeli više nisu evaluirani u odnosu na izvorne oznake događaja, nego u odnosu na umjetno promijenjene ciljne vrijednosti. Time je omogućena analiza njihove otpornosti na nelogične promjene u strukturi podataka te provjera zadržavaju li modeli konzistentnost s prethodno naučenim obrascima.
Temeljna metodološka pretpostavka ovog eksperimenta bila je da kvalitetan i stabilan model neće promijeniti logiku svog odlučivanja samo zato što su oznake glavnih događaja umjetno izmijenjene. Drugim riječima, očekivano je da modeli nastave prepoznavati stvarne odnose između ulaznih varijabli i tipova pomorskih nezgoda, čak i kada se formalne oznake događaja promijene. Takav pristup ima posebno značenje u području procjene rizika. Kada bi model bio pretjerano osjetljiv na promjene ciljnih oznaka i nekritički prihvaćao svaku novu strukturu podataka, tada bi njegova primjena u stvarnim sigurnosnim sustavima bila upitna. U kontekstu pomorske sigurnosti to bi moglo dovesti do pogrešnih procjena rizika i donošenja neadekvatnih operativnih odluka. Stoga je sposobnost očuvanja prethodno naučenih zakonitosti važan pokazatelj robusnosti modela.

Svi prethodno razvijeni modeli, uključujući model potpornih vektora, model slučajnih šuma, model neuronskih mreža, ansambl model i slojeviti model, učitani su zajedno s pripadajućim metapodacima. Slojeviti model predstavlja višerazinski pristup u kojem se izlazi osnovnih modela koriste kao ulazne značajke dodatnog modela koji uči optimalne odnose među njihovim predikcijama. Time se omogućuje dodatna integracija informacija i smanjenje utjecaja pojedinačnih pogrešaka osnovnih modela.
Prije same validacije osigurana je potpuna usklađenost ulaznih varijabli s trenirajućim skupom podataka, uključujući nazive varijabli, njihov redoslijed te razine kategorijskih atributa. Time

je omogućeno provođenje validacije u jednakim uvjetima u kojima su modeli prethodno trenirani i optimizirani.
Na temelju tako pripremljenih podataka generirane su predikcije klasa i raspodjele vjerojatnosti za svih pet modela. Ansambl model temelji se na većinskom odlučivanju između modela potpornih vektora, slučajnih šuma i neuronskih mreža, dok slojeviti model koristi njihove izlaze kao ulazne značajke i prelazi s jednostavne agregacije na učenje međusobnih odnosa među modelima.

Uspješnost validacije procijenjena je korištenjem matrica zabune i niza klasifikacijskih metrika, uključujući točnost klasifikacije, osjetljivost, specifičnost, preciznost, F1-mjeru te pokazatelje pozitivnog i negativnog omjera vjerodostojnosti. Međutim, zbog namjerne promjene ciljnih oznaka, dobivene metrike ne mogu se interpretirati na isti način kao u standardnim klasifikacijskim problemima. U ovom slučaju njihov osnovni smisao nije procjena sposobnosti modela da reproduciraju umjetno promijenjene oznake, nego procjena njihove stabilnosti i otpornosti na narušene podatke.

Posebno je važno naglasiti povezanost ove validacije s procjenom rizika, koja predstavlja središnji doprinos ovog istraživanja. Procjena rizika ne temelji se isključivo na jednoj klasifikacijskoj odluci, nego na raspodjeli vjerojatnosti po svim mogućim ishodima. Upravo zbog toga očuvanje logičnih odnosa među klasama ima veću važnost od same formalne klasifikacijske točnosti u uvjetima umjetno promijenjenih oznaka. Rezultati validacije, uključujući pojedinačne predikcije, matrice zabune i sve izračunate metrike, sustavno su pohranjeni, čime je omogućena potpuna transparentnost i ponovljivost analize. Nakon provedene validacije dobiveni su kvantitativni pokazatelji performansi za svih pet modela, čime je omogućena njihova usporedba i završna evaluacija. Rezultati prikazani u Tablici 63. potvrđuju osnovnu metodološku pretpostavku provedenog eksperimenta. Kod svih modela dolazi do izrazitog smanjenja vrijednosti metrika *accuracy*, *sensitivity* i *precision*, dok istodobno vrijednost *error rate* postaje vrlo visoka. Međutim, u kontekstu ovako postavljenog validacijskog postupka takav rezultat predstavlja očekivanu i metodološki opravdanu pojavu.

Tablica 63. Metrike modela u odnosu na kontrolirano promijenjene ciljne oznake

Model	SVM	RF	NN	Ansambl	Slojeviti
Accuracy	0,10526	0,10526	0,10526	0,10526	0,10526
ErrorRate	0,89474	0,89474	0,89474	0,89474	0,89474
Sensitivity	0,14286	0,14286	0,14286	0,14286	0,14286
Specificity	0,84853	0,84853	0,84853	0,84853	0,84912
Precision	0,14286	0,14286	0,14286	0,14286	0,14286
NPV	0,84853	0,84853	0,84853	0,84853	0,84884
F1	0,14286	0,14286	0,14286	0,14286	0,14286
FDR	0,85714	0,85714	0,85714	0,85714	0,85714
FOR	0,15147	0,15147	0,15147	0,15147	0,15116
FPR	0,15147	0,15147	0,15147	0,15147	0,15088
FNR	0,85714	0,85714	0,85714	0,85714	0,85714
LR_Positive	0,94313	0,94313	0,94313	0,94313	0,94686
LR_Negative	1,01015	1,01015	1,01015	1,01015	1,00944
DOR	0,93366	0,93366	0,93366	0,93366	0,93800

Izvor: izradio autor

Modeli su tijekom procesa treniranja usvojili stvarne odnose između ulaznih varijabli i glavnih događaja pomorskih nezgoda. Nakon što su oznake glavnih događaja namjerno promijenjene, modeli nisu promijenili logiku svog odlučivanja, nego su nastavili prepoznavati obrasce koji odgovaraju izvornoj strukturi podataka. Upravo zbog toga dolazi do formalnog smanjenja klasifikacijske točnosti u odnosu na umjetno izmijenjene ciljne vrijednosti.
Dobivena vrijednost *accuracy* od približno 10 % pokazuje da modeli nisu prilagodili svoje ponašanje umjetno promijenjenim oznakama klasa. U okviru ovakvog eksperimentalnog scenarija takvo ponašanje može se smatrati pokazateljem stabilnosti modela i očuvanja prethodno naučenih zakonitosti sustava.
Ovakav rezultat posebno je značajan u kontekstu procjene rizika. Da su modeli nakon promjene oznaka ostvarili visoku klasifikacijsku točnost, to bi moglo ukazivati na pretjeranu prilagodljivost umjetno izmijenjenim ciljnim vrijednostima i slabu povezanost s izvornom strukturom podataka. U sigurnosno kritičnim sustavima takvo bi ponašanje predstavljalo ozbiljan metodološki nedostatak jer bi upućivalo na nestabilnost modela i njihovu osjetljivost na nelogične promjene u podacima.
Vrlo slične vrijednosti metrika kod svih pet modela potvrđuju visoku razinu konzistentnosti cjelokupnog metodološkog okvira. Svi modeli reagiraju gotovo identično, što pokazuje da umjetna promjena ciljnih oznaka nije dovela do potpunog narušavanja prethodno naučenih odnosa među varijablama. Slojeviti model pritom ostvaruje neznatno bolje vrijednosti specifičnosti i negativne prediktivne vrijednosti, što ukazuje na nešto stabilnije ponašanje u uvjetima umjetno modificiranih podataka.
Dodatno, relativno visoke vrijednosti specifičnosti i negativne prediktivne vrijednosti pokazuju da modeli i dalje uspijevaju razlikovati obrasce koji ne pripadaju pojedinim klasama događaja. To upućuje na očuvanje unutarnje logike razdvajanja među kategorijama pomorskih nezgoda čak i nakon promjene ciljnih oznaka.
Vrijednosti pokazatelja *LR_Positive*, *LR_Negative* i *DOR* dodatno potvrđuju da modeli u odnosu na umjetno promijenjene oznake više nemaju stvarnu diskriminacijsku sposobnost, što je u ovom eksperimentu očekivan rezultat. Budući da cilj validacijskog postupka nije bio postizanje visoke klasifikacijske točnosti nad namjerno narušenim podacima, nego provjera zadržavaju li modeli prethodno naučene zakonitosti sustava, ovakvi rezultati predstavljaju metodološki logičan ishod.

U kontekstu procjene rizika ovi rezultati imaju posebno značenje jer upućuju na stabilnost raspodjele vjerojatnosti po klasama događaja. Budući da se procjena rizika temelji upravo na tim raspodjelama, očuvanje logičnih odnosa među događajima važnije je od same formalne klasifikacijske točnosti u umjetno modificiranom scenariju. Stabilnost raspodjele vjerojatnosti pokazuje da modeli nisu izgubili sposobnost razlikovanja sigurnosno relevantnih obrazaca ni nakon promjene ciljnih oznaka.

Zaključno, druga razina validacije provedena je kako bi se provjerilo reagiraju li modeli u očekivanom smjeru na promjenu ciljnih oznaka te zadržavaju li pritom stabilnost svojih predikcija. Nakon prve razine validacije, koja je uključivala podešavanje hiperparametara i analizu matrica zabune, dodatna provjera provedena je umjetnom izmjenom ciljnih oznaka. Rezultati pokazuju da svi razvijeni modeli strojnog učenja zadržavaju konzistentnost s prethodno naučenim obrascima i u uvjetima promijenjene strukture ciljnih podataka. Dobiveni

nalazi upućuju na stabilnost modela i očuvanje njihove unutarnje logike odlučivanja, što predstavlja važnu pretpostavku za njihovu primjenu u kvantitativnoj procjeni rizika pomorskih nezgoda i razvoju sustava potpore odlučivanju u području sigurnosti obalne plovidbe. Ova provjera ne predstavlja procjenu prediktivne uspješnosti na neovisnim, neizmijenjenim podacima.

5. RAZMATRANJA

Istraživanje je bilo usmjereno na obalnu plovidbu jer upravo taj segment pomorskog prometa generira najveći broj sigurnosno relevantnih događaja i predstavlja najkompleksnije operativno okruženje. Kombinacija ograničenog manevarskog prostora, visoke gustoće prometa, interakcije s obalom i promjenjivih meteoroloških uvjeta rezultira povećanom razinom rizika u odnosu na otvoreno more. Takav kontekst zahtijeva preciznije i dinamičnije metode procjene rizika od onih koje se tradicionalno koriste.

U ovom radu razvijeni su modeli procjene rizika koji integriraju klasifikacijske modele strojnog učenja s kvantifikacijom rizika temeljenom na vjerojatnostima i težinama posljedica. Na ovaj način je omogućen prijelaz s retrospektivne analize nezgoda na prediktivni pristup koji omogućuje procjenu rizika za konkretne situacije. Validacija na neviđenim podacima dodatno potvrđuje da modeli ne funkcioniraju samo u kontroliranim uvjetima, već zadržavaju stabilnost i u realnijim scenarijima.

Najvažnija spoznaja koja proizlazi iz rezultata jest da ukupna točnost modela nije dovoljna za ocjenu njegove vrijednosti u kontekstu sigurnosti plovidbe. Ključno je kako modeli raspoređuju pogreške između klasa. Model koji podcjenjuje rijetke, ali opasne događaje može imati visoku točnost, ali nisku operativnu vrijednost. Nasuprot tome, model koji održava ravnotežu između osjetljivosti i preciznosti daje realniju procjenu rizika i veću sigurnosnu korist. Upravo ta ravnoteža potvrđena je kod kombiniranih modela, osobito kod slojevitog modela.

5.1. Tumačenje rezultata i usporedba modela

Rezultati provedenog istraživanja pokazali su da je problem procjene rizika pomorskih nezgoda u obalnoj plovidbi moguće uspješno obraditi metodama strojnog učenja, ali istodobno su pokazali da kvaliteta modela ne ovisi samo o ukupnoj klasifikacijskoj točnosti. Za razliku od pristupa koji bi pomorski događaj promatrali samo kao pojavu ili izostanak nezgode, u ovom istraživanju problem je postavljen kao višeklasni. Promatrano je sedam kategorija glavnih događaja, pri čemu se za svaki zapis ne određuje samo pripadnost jednoj klasi, nego se procjenjuje i raspodjela vjerojatnosti po mogućim ishodima. Na temelju procijenjenih vjerojatnosti pojedinih kategorija glavnih događaja i pripadajućih razina ozbiljnosti omogućeno je daljnje određivanje razine rizika. Ovakav pristup predstavlja važan korak od same klasifikacije događaja prema njegovoj interpretaciji u kontekstu upravljanja pomorskim rizikom. Metodološki okvir ne završava predikcijom klase, nego povezuje predikciju, posljedice i rizik. U istraživanju su uspoređeni modeli potpornih vektora, slučajnih šuma i neuronskih mreža, a njihovi su rezultati dodatno povezani kroz ansambl i slojeviti pristup. Prikazana vrijednost DOR-a sažima sedam usporedbi pojedine klase sa svim ostalim klasama; razlike između modela treba tumačiti zajedno s matricama zabune i klasnim metrikama.

Usporedba osnovnih modela pokazala je očekivane razlike u njihovu ponašanju. Model potpornih vektora ostvario je točnost od 0,70763, osjetljivost od 0,66335, specifičnost od 0,94971, preciznost od 0,65891 i F1-mjeru od 0,65636. Dobivene vrijednosti su pokazale da model može razmjerno dobro razlikovati promatrane klase, ali da dio složenijih obrazaca ostaje nedovoljno razdvojen. Posebno je značajno da visoka specifičnost pokazuje dobru sposobnost

isključivanja slučajeva koji ne pripadaju promatranoj kategoriji, dok niža osjetljivost ukazuje na određeni broj slučajeva koje model ne prepoznaje dovoljno uspješno. To se vidi i u matrici zabune, gdje se pogreške češće pojavljuju među klasama koje dijele slične operativne i okolišne karakteristike. Takvo ponašanje je očekivano kod višeklasnog problema u kojem granice između pojedinih klasa pomorskih događaja nisu uvijek jasno odvojene.

Model slučajnih šuma pokazao je najbolju ukupnu klasifikacijsku uspješnost među pojedinačnim modelima. Njegova točnost iznosi 0,76695, osjetljivost 0,79694, specifičnost 0,96000, preciznost 0,71554 i F1-mjera 0,73488, dok DOR iznosi 200,44876. Ovakva kombinacija pokazatelja posebno je značajna u procjeni pomorskog rizika jer pokazuje da model uspješno prepoznaje velik udio relevantnih slučajeva uz istodobno visoku sposobnost razlikovanja slučajeva koji ne pripadaju promatranoj klasi. Visoka vrijednost osjetljivosti ima posebno značenje u sigurnosnom kontekstu. Propuštanje događaja koji pripada visoko relevantnoj klasi može imati ozbiljnije posljedice od njihova pogrešnog označavanja u manje relevantnu kategoriju. Istodobno, visoka specifičnost ograničava broj slučajeva koje bi model nepotrebno označio kao pripadajuće promatranoj klasi. Upravo takav odnos između osjetljivosti i specifičnosti daje modelu slučajnih šuma praktičnu vrijednost za daljnju primjenu u procjeni rizika. Rezultati, međutim, ne znače da model jednako uspješno prepoznaje svih sedam glavnih događaja. Rijetke i međusobno slične klase ostaju zahtjevnije za klasifikaciju, što pokazuje da ukupna vrijednost modela može prikriti razlike između pojedinih kategorija.

Rezultati modela neuronske mreže dodatno potvrđuju da složenost odnosa među varijablama ima važnu ulogu u promatranom problemu. Neuronska mreža ostvarila je točnost od 0,71610, osjetljivost 0,68854, specifičnost 0,95094, preciznost 0,65614, F1-mjeru 0,66822 i DOR 57,38777. Njezini rezultati su potvrdili mogućnost modeliranja nelinearnih odnosa, ali istodobno su pokazali veću osjetljivost na konfiguraciju modela i strukturu podataka. Tijekom razvoja bilo je potrebno optimizirati broj neurona u skrivenom sloju i regularizacijski parametar te dodatno reducirati ulazne varijable kako bi se postigla stabilnija konvergencija. Konačno je odabrana konfiguracija sa jednim skrivenim slojem, sa šest neurona te *decay values* = 0,5.

Razlike između osnovnih modela stoga nisu samo pitanje toga koji je model „bolji“, nego pokazuju da algoritmi na različite načine prepoznaju istu strukturu podataka. Potporni vektori, slučajne šume i neuronske mreže koriste različite mehanizme za formiranje klasifikacijske odluke i zbog toga različito reagiraju na preklapanje klasa, nelinearnost i međudjelovanje varijabli. U takvom problemu usporedba više modela je dala potpuniju sliku nego odabir samo jednog algoritma unaprijed.

Ansambl model pokazao je da kombiniranje osnovnih modela može dati stabilnu i uravnoteženu izvedbu. Ostvario je točnost 0,73729, osjetljivost 0,73306, specifičnost 0,95485, preciznost 0,67652, F1-mjeru 0,68761 i DOR 82,27012. Iako nije nadmašio slučajne šume prema ukupnoj točnosti, zadržao je visoku specifičnost i relativno uravnotežene vrijednosti ostalih pokazatelja. Njegova vrijednost stoga nije u tome da je prema svakoj metrici najbolji model, nego u tome što smanjuje ovisnost konačne odluke o jednom algoritmu i koristi prednosti različitih modelskih pristupa.

Slojeviti model ostvario je točnost 0,73305, osjetljivost 0,72307, specifičnost 0,95394, preciznost 0,67025, F1-mjeru 0,68602 i DOR 101,15316. Dakle, ni on nije nadmašio slučajne šume u ukupnoj točnosti. Njegova vrijednost proizlazi iz toga da se predikcije osnovnih modela

ne agregiraju samo većinskim odlučivanjem, nego se koriste kao ulaz meta-modelu koji uči odnose među njihovim predikcijama. Time slojeviti model predstavlja metodološki napredniji oblik integracije, ali ne nužno model s najvećom pojedinačnom klasifikacijskom točnošću. Ovakav nalaz je važan i za interpretaciju njegove primjene u procjeni rizika.

U ovom istraživanju cilj nije bio samo odabrati model s najvećom točnošću, nego dobiti što korisniju raspodjelu vjerojatnosti koja se potom povezuje s različitim težinama posljedica. Također nije moguće zaključiti da složeniji model automatski daje bolji rezultat. U konkretnom istraživanju model slučajnih šuma je pokazao najbolju ukupnu prediktivnu uspješnost među uspoređenim pristupima, dok su ansambl i slojeviti modeli pokazali vrijednost u pogledu stabilnosti i kombiniranja različitih modelskih perspektiva. Time je potvrđeno da izbor modela treba biti povezan s ciljem primjene i vrstom pogreške koja se želi minimizirati, a ne samo s nastojanjem da se postigne najveća moguća vrijednost jedne metrike.

Posebno je značajno da su vjerojatnosti dobivene iz modela povezane s razinama ozbiljnosti događaja, a zatim i s tri razine rizika. Time je ostvarena poveznica između klasifikacije i stvarne procjene sigurnosnog značaja događaja. Događaj koji ima relativno visoku vjerojatnost, ali manje ozbiljne posljedice, ne mora imati isti prioritet kao manje vjerojatan događaj s vrlo teškim mogućim posljedicama. Uključivanjem posljedica u konačni rezultat dobiva se realniji prikaz rizika nego kada bi se model promatrao samo kroz učestalost ili klasifikacijsku točnost. Upravo se u ovom dijelu nalazi jedna od važnijih spoznaja istraživanja a to je da klasifikacijska uspješnost i procijenjena razina rizika nisu ista veličina. Model s najvećom točnošću ne mora nužno generirati najveću procjenu rizika, jer konačni rezultat ovisi i o raspodjeli vjerojatnosti po klasama i pripadajućim indeksima posljedica. Time se opravdava odvojeno promatranje performansi modela i izvedene procjene rizika. Ovim se složeni izlaz višeklasnog modela pretvara u informaciju koja je razumljiva za potrebe upravljanja rizikom.

Ovakav pristup posebno je važan jer učestalost pomorskih događaja sama po sebi ne predstavlja potpunu mjeru rizika. Dva područja ili dva tipa događaja mogu imati sličnu učestalost, ali različite moguće posljedice. Uključivanjem ozbiljnosti posljedica omogućuje se razlikovanje događaja prema njihovu sigurnosnom značaju. Na taj način model ne daje samo informaciju o tome koji se događaj najvjerojatnije može pojaviti, nego i informaciju o tome koliki sigurnosni značaj taj događaj može imati.

5.2. Značaj najvažnijih varijabli za pomorski rizik

Analiza važnosti varijabli je pokazala da se pomorski rizik u analiziranom skupu podataka oblikuje kroz kombinaciju više skupina čimbenika. U ulaznom skupu bilo je 29 varijabli koje obuhvaćaju karakteristike broda, vremenski i analitički kontekst pomorske nezgode, okolišni kontekst, operativne značajke te posljedice.

Kod modela slučajnih šuma posebno su se izdvojili *stanje mora*, *jačina vjetra* i *ozbiljnost nezgode*, zatim *mjesec, oštećenje, tip broda, peljar, zastava broda, segment putovanja* i *područje nezgode*. Takva raspodjela važnosti pokazuje da model ne oslanja svoju klasifikaciju na jednu skupinu varijabli, nego povezuje okolišne, tehničke i operativne karakteristike. *Jačina vjetra* i *stanje mora* mogu utjecati na manevarske sposobnosti plovila, mogućnost održavanja planiranog kursa i brzinu razvoja neželjenog događaja. U obalnoj

plovidbi njihov značaj može biti dodatno naglašen zbog ograničenog prostora za manevriranje i blizine drugih plovila, objekata ili obale.

Rezultati modela potpornih vektora također su pokazali izražen doprinos okolišnih i prostornih varijabli, ali uz veći utjecaj pojedinih administrativnih i vremensko-analitičkih obilježja. Kod neuronske mreže istaknute su varijable *ozbiljnosti nezgode*, *jačina vjetra*, *stanja mora*, *oštećenja* i administrativnog konteksta, što potvrđuje da različiti modeli prepoznaju različite obrasce iz istog skupa podataka.

Prostorni kontekst posebno je važan za obalnu plovidbu. Područje u kojem se *glavni događaj* promatra može imati ograničen manevarski prostor, različitu gustoću prometa, različitu razinu interakcije među plovilima i različite mogućnosti izbjegavanja opasnosti. Zbog toga istovjetna tehnička karakteristika broda ne mora imati jednak sigurnosni značaj u različitim područjima plovidbe. Slično vrijedi i za meteorološke i oceanografske uvjete. *Stanje mora, jačina vjetra* i *vidljivost* međusobno su povezani čimbenici koji mogu utjecati na sigurnost plovidbe. Njihov istodobni utjecaj može promijeniti mogućnost održavanja kursa, manevarsku sposobnost, preglednost navigacijskog prostora i raspoloživo vrijeme za reakciju. Rezultati korelacijske analize u poglavlju 4. dodatno su pokazali izraženu povezanost između *stanja mora* i *jačine vjetra* te između *stanja mora* i *vidljivosti*, što podupire interpretaciju njihovih rezultata u modelima. Važan je i doprinos karakteristike plovila. *Tip broda, kategorija, starost, zastava, dimenzije, posada i prisutnost peljara* opisuju različite dijelove operativnog sustava. Njihov utjecaj treba tumačiti zajedno s okolišnim i prostornim čimbenicima, jer se stvarni pomorski događaj razvija upravo kroz njihovu kombinaciju.

Za praktičnu primjenu posebno je važna interpretabilnost modela. Prediktivna sposobnost sama po sebi nije dovoljna ako korisnik sustava ne može razumjeti koji čimbenici najviše doprinose procijenjenom riziku. Analiza važnosti varijabli stoga predstavlja važan dio istraživanja jer omogućuje povezivanje dobivenih predikcija s konkretnim obilježjima pomorskih događaja. Na taj način model nije samo alat za predviđanje nego može poslužiti i za prepoznavanje uvjeta koji su povezani s povećanom razinom rizika. Ipak, važno je naglasiti da važnost varijable u prediktivnom modelu ne znači nužno postojanje izravne uzročno-posljedične veze. Rezultate stoga treba tumačiti kao statističke odnose identificirane u dostupnim podacima, a ne kao dokaz da promjena pojedine varijable sama po sebi uzrokuje promjenu rizika.

5.3. Doprinos istraživanja

Doprinos istraživanja može se precizno sagledati na tri razine: analitičkoj, metodološkoj i operativnoj. Na analitičkoj razini, rad pokazuje da se sigurnosni događaji u obalnoj plovidbi mogu modelirati kao višeklasni problem s jasnom raspodjelom vjerojatnosti, pri čemu svaki ulazni zapis ne pripada jednoj kategoriji deterministički, već s određenim stupnjem vjerojatnosti više kategorija. Takav pristup omogućuje realnije sagledavanje nesigurnosti i prelazak s klasifikacije na procjenu rizika.

Na razini rezultata jasno je pokazano da pojedinačni modeli imaju različite prednosti i ograničenja. Model potpornih vektora daje stabilne granice između klasa, ali je manje osjetljiv na složene interakcije. Model slučajnih šuma učinkovito identificira nelinearne odnose i važnost ulaznih varijabli, dok neuronske mreže omogućuje modeliranje složenih nelinearnih

obrazaca, ali mogu biti osjetljivije na konfiguraciju modela i strukturu podataka. Razlike između rezultata pojedinih modela stoga ne treba promatrati isključivo kao nedostatak, nego kao pokazatelj da različiti algoritmi na različite načine prepoznaju strukturu promatranog problema.

Doprinos usporedbe modela jest prikaz različitih načina klasifikacije i kombiniranja njihovih predikcija. Najveću ukupnu točnost ostvario je model slučajnih šuma (0,767), dok su ansambl (0,737) i slojeviti model (0,733) ostvarili niže vrijednosti. Kombinirani pristupi stoga u ovom istraživanju nisu pokazali prednost prema ukupnoj točnosti, ali omogućuju analizu učinka povezivanja osnovnih modela na klasifikaciju pojedinih vrsta događaja.

Metodološki doprinos rada ogleda se u razvoju cjelovitog postupka koji povezuje klasifikaciju pomorskih događaja, vjerojatnosti i konačno izračun tri razine rizika. Uveden je sustav evaluacije koji uključuje veći broj metrika uključujući točnost, osjetljivost, specifičnost, preciznost, F1-mjeru i DOR, prilagođenih višeklasnom problemu, čime je omogućena potpunija analiza ponašanja modela od one koja bi se temeljila samo na ukupnoj točnosti. Time se izlaz modela pretvara iz same klasifikacijske vrijednosti u informaciju koja ima izravnije značenje za procjenu i upravljanje pomorskim rizikom.

Posebno važan metodološki doprinos predstavlja i validacija modela na neviđenim podacima. Takav pristup omogućuje procjenu sposobnosti modela da zadrži prediktivnu kvalitetu na podacima koji nisu sudjelovali u njegovoj izgradnji te daje realniju procjenu njegove generalizacijske sposobnosti. Ovim pristupom se smanjuje mogućnost donošenja zaključaka koji bi bili posljedica isključivo prilagodbe modela podacima korištenima tijekom razvoja.

Doprinos istraživanja dodatno se ogleda u uključivanju prostornog i prometnog konteksta u procjenu rizika. Uvažavanjem kategorije područja u kojem je došlo do neželjenog događaja, gustoće prometa i karakteristika plovila, rizik se ne promatra kao posljedica jednog izoliranog čimbenika, već kao rezultat međudjelovanja različitih karakteristika pomorskog prometnog sustava. Ovim se stvara metodološka osnova za daljnje povezivanje prediktivnog modela s prostornim informacijama i geografskim informacijskim sustavima.

Operativni doprinos istraživanja očituje se u mogućnosti primjene razvijenog modela kao osnove za buduće sustave potpore odlučivanju u području pomorske sigurnosti. Model omogućuju da se za svaki konkretan slučaj izračuna razina rizika, što može služiti kao podrška pri donošenju odluka, planiranju plovidbe i upravljanju sigurnošću. Njegova operativna uporabljivost i učinak na sigurnost plovidbe još nisu ispitani.

Dodatni operativni potencijal proizlazi iz mogućnosti integracije razvijenog modela s relevantnim izvorima pomorskih podataka, uključujući Europsku mrežu za nadzor i prikupljanje podataka o moru (engl. *European Marine Observation and Data Network* – EMODnet), EMSA-u i GISIS, AIS podacima kao i s drugim prostornim, prometnim, meteorološkim i oceanografskim izvorima podataka. Takva integracija omogućila bi razvoj sustava koji bi povijesne podatke o pomorskim događajima koristio za prepoznavanje obrazaca i područja povećanog rizika te pružao dodatnu kvantitativnu informaciju korisnicima uključenima u upravljanje pomorskom sigurnošću.

Poseban doprinos predstavlja kombiniranje klasifikacije s procjenom ozbiljnosti i rizika. Sedam glavnih događaja povezano je s četiri razine ozbiljnosti, a zatim s tri razine rizika.

Time je izlaz modela pretvoren iz same statističke predikcije u informaciju koja može imati praktično značenje za upravljanje sigurnošću.

Znanstvena vrijednost istraživanja stoga nije samo u postizanju određene razine prediktivne uspješnosti, već u povezivanju više elemenata u jedinstven metodološki okvir: podataka o pomorskim događajima, metoda strojnog učenja, procjene ozbiljnosti posljedica, razina rizika i prostornog konteksta. Ovakav pristup omogućuje prijelaz od tradicionalne analize već nastalih pomorskih nezgoda prema proaktivnijem pristupu u kojem se postojeći podaci koriste za prepoznavanje uvjeta povezanih s povećanim rizikom.

Razvijeni model pritom ne treba promatrati kao zamjenu za stručnu prosudbu zapovjednika, peljara, operatora ili nadležnih pomorskih tijela, već kao alat potpore odlučivanju koji pruža dodatnu kvantitativnu informaciju o pripadnosti evidentiranih događaja pojedinim klasama i njihovoj povezanoj razini posljedica. Njegova praktična vrijednost proizlazi iz mogućnosti povezivanja analitičkih rezultata sa stručnom prosudbom i procesima upravljanja sigurnošću te usmjeravanja preventivnih mjera prema vrstama događaja i obilježjima koja su u analiziranom skupu povezana s višom razinom posljedica.

5.4. Ograničenja istraživanja

Unatoč značajnim rezultatima, istraživanje ima nekoliko jasno identificiranih ograničenja koja je potrebno uzeti u obzir pri interpretaciji i mogućoj praktičnoj primjeni modela.

Prvo ograničenje odnosi se na podatke. Modeli su trenirani na dostupnim podacima, što znači da kvaliteta i obuhvat podataka izravno određuju kvalitetu rezultata. Podaci o evidentiranim nezgodama predstavljaju događaje koji su se već dogodili te stoga ne obuhvaćaju nužno sve opasne situacije koje nisu rezultirale prijavljenom nezgodom. Uključivanje podataka o *near-miss* događajima moglo bi povećati sposobnost modela za proaktivnu procjenu rizika, jer bi se osim stvarno nastalih nezgoda mogli analizirati i događaji kod kojih je nezgoda izbjegnuta. Također nedostatak određenih varijabli, poput detaljnih meteoroloških uvjeta ili podataka u stvarnom vremenu, može ograničiti sposobnost modela da prepozna određene rizične situacije. Budući da ljudska pogreška nije zasebno evidentirana, a nisu dostupni ni podaci o opsegu plovidbe po vremenskim uvjetima, ovim istraživanjem nije bilo moguće provjeriti hipotezu o njihovoj kombinaciji kao najčešćem uzroku nezgoda. Dodatno ograničenje predstavlja velik udio nedostajućih vrijednosti za pojedine prediktore, osobito vremenske uvjete, vidljivost, broj članova posade, stanje mora i jačinu vjetra. Imputacija omogućuje obradu zapisa, ali ne nadomješta izvorno opažene podatke; stoga rezultate koji ovise o tim varijablama treba tumačiti s oprezom. S obzirom na to da je evidentirana ozbiljnost događaja korištena i pri klasifikaciji i pri izvođenju indeksa posljedica, te dvije komponente nisu neovisne, pa tako dobiven rezultat ne predstavlja prospektivnu procjenu rizika prije nastanka događaja.

Drugo ograničenje je neuravnoteženost klasa. Rijetki događaji, koji su često i najopasniji, imaju manji broj primjera u podacima. Ovo otežava njihovo učenje i može dovesti do njihove podzastupljenosti u predikcijama, što ima izravan utjecaj na procjenu rizika.

Treće ograničenje odnosi se na složenost modela. Budući da su kombinirani modeli složeniji, njihova interpretacija je složenija. U operativnim uvjetima to može predstavljati problem jer korisnici trebaju razumjeti razloge predikcije, a ne samo rezultat.

Četvrto ograničenje vezano je uz definiranje težina posljedica u izračunu rizika. Težine su temeljene na stručnim procjenama i standardima, ali nisu potpuno objektivne. Promjena tih vrijednosti može utjecati na konačne rezultate procjene rizika.

Peto ograničenje odnosi se na područje primjene. Model je razvijen za obalnu plovidbu i njegova primjena u drugim uvjetima zahtijeva dodatnu validaciju. Također, iako je provedena validacija na neviđenim podacima, model nije implementiran u stvarni operativni sustav, što je nužno za potpunu potvrdu njegove vrijednosti. Struktura pomorskog prometa, vrste plovila, uvjeti plovidbe, način izvještavanja i regulatorni okvir mogu se razlikovati između različitih područja. Zbog toga rezultate nije opravdano automatski generalizirati na sva pomorska područja bez dodatne vanjske validacije. Buduća istraživanja trebala bi uključiti podatke iz drugih geografskih područja kako bi se provjerila stabilnost dobivenih odnosa i sposobnost modela da zadrži prediktivnu kvalitetu izvan skupa na kojem je razvijen.

Dodatno ograničenje predstavlja agregiranje izvornih kategorija događaja u četiri razine ozbiljnosti i njihovo naknadno grupiranje u tri razine rizika. Iako se time gubi dio detalja sadržanog u izvornim klasama, takav pristup omogućuje stvaranje rezultata koji je razumljiviji i primjenjiviji u operativnom sustavu procjene rizika. U tom smislu agregacija predstavlja kompromis između detaljnosti podataka i praktične uporabljivosti konačnog rezultata.

Rezultati testiranja na neviđenim podacima potvrđuju prediktivnu sposobnost u uvjetima istraživanja, ali puna potvrda operativne vrijednosti zahtijevala bi pilot-implementaciju i dugotrajnije testiranje u stvarnim uvjetima pomorske plovidbe.

5.5. Preporuke i smjernice za daljnja istraživanja

Rezultati istraživanja jasno ukazuju na konkretne smjerove daljnjeg razvoja modela i njegove primjene u procjeni rizika obalne plovidbe.

Prvi smjer odnosi se na unapređenje podataka. Uključivanje meteoroloških podataka visoke rezolucije, podataka o stanju mora, gustoći prometa i kretanju plovila u stvarnom vremenu omogućilo bi preciznije i dinamičnije procjene rizika. Integracija AIS podataka s meteorološkim i oceanografskim informacijama posebno bi omogućila razvoj prostorno-vremenskog modela koji bi procjenjivao rizik ne samo na temelju povijesnih obilježja nego i prema trenutačnim uvjetima plovidbe.

Drugi smjer je razvoj prostorno-vremenskih modela. Rizik u obalnoj plovidbi varira ovisno o lokaciji i vremenu, pa je potrebno razviti modele koji mogu identificirati rizična područja i vremenska razdoblja povećane opasnosti.

Treći smjer je unapređenje rada s neuravnoteženim podacima. Primjenom naprednih metoda moguće je poboljšati prepoznavanje rijetkih, ali sigurnosno značajnih događaja, čime bi se dodatno povećala praktična vrijednost modela za procjenu pomorskog rizika.

Četvrti smjer odnosi se na interpretabilnost modela. Primjena metoda objašnjive umjetne inteligencije omogućila bi bolje razumijevanje čimbenika koji utječu na pojedine

predikcije, povećala transparentnost modela i olakšala njegovo prihvaćanje od strane krajnjih korisnika.

Peti smjer, koji je ujedno i najvažniji, jest testiranje modela u stvarnim uvjetima. Implementacija u pilot-sustave nadzora plovidbe omogućila bi provjeru njegove operativne vrijednosti i dodatnu optimizaciju.

Posebno važan dio budućih istraživanja predstavlja proširena validacija modela. Druga razina validacije provedena je kontroliranom izmjenom ciljnih oznaka što predstavlja dodatnu provjeru ponašanja modela u uvjetima narušene strukture podataka, ali je treba promatrati kao eksperimentalni scenarij, a ne kao konačan dokaz robusnosti prema svim mogućim oblicima promjene podataka. Buduća istraživanja trebala bi stoga uključiti veći broj zapisa i različite obrasce narušavanja podataka.

6. ZAKLJUČAK

Polazeći od problema nedostatne učinkovitosti tradicionalnih metoda procjene rizika u uvjetima složenog i dinamičnog okruženja obalne plovidbe, u ovom radu razvijen je model procjene rizika temeljen na metodama strojnog učenja koji omogućuje analizu, klasifikaciju i predikciju pomorskih nezgoda i procjenu rizika. Time je problem istraživanja izravno riješen kroz konstrukciju analitičkog sustava koji je sposoban kvantificirati rizik na temelju stvarnih podataka i međudjelovanja čimbenika koji na njega utječu.

Predmet istraživanja operacionaliziran je kroz primjenu modela potpornih vektora, slučajnih šuma i neuronskih mreža, kao i njihovih kombinacija putem ansambl i slojevitog pristupa. Modeli su razvijeni na temelju empirijskih podataka koji obuhvaćaju karakteristike plovila, operativne i okolišne varijable, vremenski kontekst te pokazatelje posljedica pomorskih događaja, čime je omogućeno modeliranje stvarnih uvjeta obalne plovidbe. Ovim je ostvarena temeljna funkcija modela, klasifikacija sedam klasa pomorskih događaja uz istodobno određivanje njihove vjerojatnosti.

Istraživanjem je razvijen postupak klasifikacije sedam vrsta glavnih pomorskih događaja i povezivanja procijenjenih vjerojatnosti pripadnosti tim klasama s empirijskim indeksima posljedica. Tako dobiveni pokazatelj opisuje relativnu težinu mogućih posljedica unutar skupa evidentiranih događaja. Budući da istraživanje ne obuhvaća plovidbene situacije bez događaja ni mjeru izloženosti, njime nije procijenjena apsolutna vjerojatnost nastanka nezgode na određenoj ruti ili tijekom određene plovidbe.

Analiza važnosti varijabli pokazala je da procjena rizika proizlazi iz međudjelovanja različitih skupina čimbenika. Posebno su se istaknuli obilježja ozbiljnosti događaja, meteorološki i oceanografski uvjeti, prostorni kontekst, karakteristike plovila i operativne karakteristike plovidbe. Takav rezultat potvrđuje da pomorski rizik nije moguće adekvatno opisati jednim čimbenikom, nego zahtijeva integrirani pristup koji uzima u obzir različite dimenzije pomorskog prometnog sustava.

Ciljevi istraživanja ostvareni su kroz identificiranje dominantnih čimbenika rizika, razvoj i validaciju modela te uspostavu postupka kojim se klasifikacijski rezultati izravno prevode u procjenu rizika. Posebno je značajno da je validacija provedena na neviđenim podacima, čime je potvrđena sposobnost modela da zadrže stabilne performanse izvan trening skupa, odnosno njihova generalizacijska sposobnost.

Glavna hipoteza potvrđena je u cijelosti. Rezultati istraživanja podupiru glavnu hipotezu u okviru razvijenog metodološkog pristupa, budući da modeli strojnog učenja omogućuju točniju i pouzdaniju procjenu rizika u odnosu na tradicionalne metode, ne samo kroz poboljšanje klasifikacijskih rezultata, već kroz sposobnost modeliranja nelinearnih odnosa i generiranja stabilnih vjerojatnosnih procjena. Upravo ova sposobnost predstavlja ključnu prednost u uvjetima obalne plovidbe gdje rizik nastaje kao rezultat međudjelovanja više čimbenika. Tradicionalna procjena pomorskog rizika ne predstavlja jedan jedinstveni algoritam koji se može jednostavno uključiti kao šesti model uz slučajne šume, potporne vektore, neuronsku mrežu, ansambl i slojeviti model. Potvrđivanje glavne hipoteze u doktorskom radu temelji se na dvije povezane razine. Prva je teorijsko-komparativna, u kojoj je kroz pregled postojeće literature analizirana sposobnost tradicionalnih i suvremenih metoda u radu sa složenim, višedimenzionalnim i nelinearnim sigurnosnim problemima. Druga je vlastita

empirijska razina, u kojoj je pokazano da razvijeni modeli mogu provesti višeklasnu klasifikaciju, prepoznati nelinearne odnose, odrediti važnost varijabli, generirati raspodjele vjerojatnosti i integrirati ih u kvantitativnu procjenu rizika. Stoga se potvrda glavne hipoteze ne temelji na tvrdnji da je izveden jedan izolirani eksperiment „strojno učenje protiv jedne tradicionalne metode“, nego na ukupnosti teorijske i empirijske argumentacije**.** Rezultati jasno pokazuju da pojedinačni modeli imaju različite karakteristike, ali nijedan ne daje optimalne rezultate u svim klasama. Model potpornih vektora pokazuje stabilnost u razdvajanju klasa, model slučajnih šuma učinkovit je u prepoznavanju interakcija varijabli, dok neuronske mreže omogućuju detekciju složenih obrazaca. Usporedba je pokazala da kombinirani modeli proširuju metodološke mogućnosti klasifikacije, ali prema ukupnoj točnosti nisu nadmašili model slučajnih šuma. Ansambl i slojeviti model predstavljaju dva načina povezivanja predikcija osnovnih modela. Završna provjera s namjerno promijenjenim oznakama nije pokazala prednost slojevitog modela prema ukupnoj točnosti niti se može koristiti kao dokaz njegove veće generalizacijske sposobnosti.

Prvu pomoćnu hipotezu, prema kojoj su najčešći uzroci pomorskih nezgoda kombinacija ljudske pogreške i nepovoljnih vremenskih uvjeta, nije bilo moguće empirijski potvrditi na temelju korištenog skupa podataka. Ljudska pogreška nije bila zasebno evidentirana kao varijabla, a raspodjela vremenskih uvjeta među evidentiranim događajima ne omogućuje utvrđivanje njihove uzročne uloge ni usporedbu rizika između različitih vremenskih uvjeta. Provjera te hipoteze zahtijevala bi pouzdane podatke o ljudskom čimbeniku i izloženosti plovidbi u pojedinim vremenskim uvjetima.

Druga pomoćna hipoteza također je potvrđena. Rezultati podupiru pomoćnu hipotezu jer razvijeni modeli daju osnovu za sposobnost pravovremenog prepoznavanja rizičnih situacija, što ih čini prikladnima za integraciju u sustave nadzora plovidbe i podrške odlučivanju. Njihova primjena može omogućiti ranije upozoravanje, kvalitetnije donošenje odluka i smanjenje učestalosti pomorskih nezgoda, dok bi stvarni učinak na smanjenje učestalosti nezgoda trebalo potvrditi operativnim testiranjem.

Praktična vrijednost rada očituje se u mogućnosti primjene modela u operativnim sustavima, gdje mogu služiti za procjenu trenutačne razine rizika i identifikaciju kritičnih čimbenika. Time se omogućuje prijelaz s reaktivnog na proaktivno upravljanje sigurnošću, što je ključan preduvjet za smanjenje broja nezgoda u obalnoj plovidbi.

Analiza važnosti varijabli pokazala je da se obrasci rizika razlikuju među vrstama pomorskih događaja. Raznolikost važnosti varijabli potvrđuje da je pomorski rizik višedimenzionalan i da je njegova procjena prikladnija kroz integrirani model nego kroz jedan izolirani indikator.

Znanstveni doprinos rada ogleda se u razvoju takvog integriranog metodološkog okvira. Povezivanjem klasifikacije, procijenjenih vjerojatnosti, ozbiljnosti posljedica i konačne razine rizika omogućeno je prevođenje složenog izlaza strojnog modela u informaciju relevantnu za upravljanje sigurnošću. Doprinos je dodatno proširen primjenom i usporedbom različitih algoritama te njihovim kombiniranjem u ansambl i slojeviti pristup.

Praktična vrijednost istraživanja očituje se u mogućnosti integracije razvijenog modela s prethodno navedenim geografskim informacijskim sustavima i relevantnim izvorima pomorskih podataka, čime se stvaraju preduvjeti za razvoj dinamičkog sustava potpore

odlučivanju. Takav sustav mogao bi omogućiti prostorno i vremenski osjetljiviju procjenu rizika te pomoći pri određivanju prioriteta za preventivno djelovanje.

Unatoč rezultatima, primjena modela ostaje ograničena kvalitetom i reprezentativnošću dostupnih podataka, neravnomjernom zastupljenošću pojedinih klasa, dostupnošću varijabli u stvarnom vremenu, načinom određivanja težina posljedica i činjenicom da model nije testiran u stvarnom operativnom sustavu. Ova ograničenja ne umanjuju ostvareni doprinos istraživanja, ali određuju granice unutar kojih je rezultate potrebno interpretirati. Budući razvoj trebao bi biti usmjeren na proširenje baze podataka uključivanjem *near-miss* događaja, AIS podataka te detaljnijih meteoroloških i oceanografskih informacija, razvoj prostorno-vremenskih modela, unapređenje rada s neuravnoteženim klasama, kalibraciju procijenjenih vjerojatnosti i primjenu metoda objašnjive umjetne inteligencije. Posebno važan korak predstavlja vanjska i vremenska validacija te pilot-implementacija u stvarnom sustavu potpore odlučivanju.

Zaključno, ovaj doktorski rad potvrđuje da primjena modela procjene rizika temeljenih na strojnome učenju predstavlja učinkovit i konkretan način povećanja sigurnosti obalne plovidbe. Razvijeni pristup omogućuje identifikaciju ključnih čimbenika rizika, predviđanje pomorskih događaja i njihovu kvantifikaciju, čime se stvara pouzdana osnova za donošenje sigurnosno utemeljenih odluka u pomorskom prometu.

7. LITERATURA

1. Agorkua, G., Hernandez, S., Falquez, M., Poddar, S., & Pang, S. (2024). Predicting barge presence and quantity on inland waterways using vessel tracking data: A machine learning approach. *arXiv:2501.00615 [cs.LG].* https://doi.org/10.48550/arXiv.2501.00615
2. Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3), 91-93. https://doi.org/10.1016/j.tjem.2018.08.001
3. Antão, P., Papanikolaou, A., Papanikolaou, D.-N., Sideris, S., Mavridou, M., & Karapidakis, G. (2024). A statistical analysis of ship accidents (1990-2020) focusing on collision, grounding, hull failure, and resulting hull damage. *Journal of Marine Science and Engineering, 12*(1), Article 122. https://doi.org/10.3390/jmse12010122
4. Atak, Ü., & Arslanoglu, Y. (2021). Machine learning methods for predicting marine port accidents: A case study in container terminal. *Ships and Offshore Structures, 16*(1), 1–8. https://doi.org/10.1080/17445302.2021.2003067
5. Australian Transport Safety Bureau (ATSB). (2025). *Marine investigation reports.* https://www.atsb.gov.au/marine-investigation-reports
6. Aven, T. (2012). *Foundations of Risk Analysis: A Knowledge and Decision-Oriented Perspective* (2nd ed.). Wiley. http://repo.darmajaya.ac.id/4601/1/Terje%20Aven%20-%20Foundations%20of%20Risk%20Analysis_%20A%20Knowledge%20and%20Decision-oriented%20Perspective-Wiley%20%282003%29.pdf.
7. Balas, E. A., & Balas, C. E. (2025). Maritime Risk Assessment: A Cutting-Edge Hybrid Model Integrating Automated Machine Learning and Deep Learning with Hydrodynamic and Monte Carlo Simulations. *Journal of Marine Science and Engineering, 13*(5), 939. https://doi.org/10.3390/jmse13050939
8. Beck, M. W. (2018). *NeuralNetTools: Visualization and Analysis Tools for Neural Networks*. *Journal of Statistical Software, 85*(11), 1–20. https://www.jstatsoft.org/article/view/v085i11
9. Bi, J., Gao, M., Zhang, W., Zhang, X., Bao, K., & Xin, Q. (2022). Research on navigation safety evaluation of coastal waters based on dynamic irregular grid. *Journal of Marine Science and Engineering, 10*(6), 733. https://doi.org/10.3390/jmse10060733
10. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
11. Brownlee, J. (2020). *How to Calculate Feature Correlation Matrix with Pandas*. Machine Learning Mastery. https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data
12. Brownlee, J. (2021). *Ensemble learning algorithms with Python*. Machine Learning Mastery. https://machinelearningmastery.com/wpcontent/uploads/2021/03/ensemble_learning_algorithms_with_python_sample.pdf
13. Bundesstelle für Seeunfalluntersuchung (BSU). (2025). *BSU official website.* https://www.bsu-bund.de/EN/Home/home_node.html

14. Bye, R. J., & Aalberg, A. L. (2018). Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports. *Reliability Engineering & System Safety, 176*, 174–186. https://doi.org/10.1016/j.ress.2018.03.033
15. Cao, Y., Iulia, M., Majumdar, A., Feng, Y., Xin, X., Wang, X., Wang, H., & Yang, Z. (2025). Investigation of the risk influential factors of maritime accidents: A novel topology and robustness analytical framework. *Reliability Engineering & System Safety, 254*(Part B), 110636. https://doi.org/10.1016/j.ress.2024.110636
16. Cao, Y., Wang, X., Wang, Y., Fan, S., Wang, H., Yang, Z., & Shi, R. (2023). Analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network. *Ocean Engineering, 269*, 113563. https://doi.org/10.1016/j.oceaneng.2022.113563
17. Chen, P., Huang, Y., Mou, J., & van Gelder, P. H. A. J. M. (2019). Probabilistic risk analysis for ship–ship collision: State-of-the-art. *Safety Science, 117*, 108–122. https://doi.org/10.1016/j.ssci.2019.04.014
18. Chen, T., Wang, H., Cai, Y., Liang, M., & Meng, Q. (2025). Exploring key factors for long-term vessel incident risk prediction. *Reliability Engineering & System Safety, 253*, 110565. https://doi.org/10.1016/j.ress.2024.110565
19. Chen, Y., Yu, Q., Wang, W., & Wu, X. (2024). Dynamic calculation approach of the collision risk in complex navigable water. *Journal of Marine Science and Engineering, 12*(9), Article 1605. https://doi.org/10.3390/jmse12091605
20. Coraddu, A., Oneto, L., Navas de Maya, B., & Kurt, R. (2020). Determining the most influential human factors in maritime accidents: A data-driven approach. *Ocean Engineering, 211*, 107588. https://doi.org/10.1016/j.oceaneng.2020.107588
21. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018
22. Danish Maritime Accident Investigation Board (DMAIB). (2025). *DMAIB official website.* https://www.dmaib.dk/
23. DHMZ. (2025). *Prognoza i podaci - Državni hidrometeorološki zavod Republike Hrvatske*. https://meteo.hr
24. Dietterich, T. G. (2000). Ensemble methods in machine learning. *Multiple Classifier Systems*, 1–15. https://doi.org/10.1007/3-540-45014-9_1
25. Du, Z. (2024). A risk assessment model for navigation safety of maritime aquaculture platform based on AIS ship trajectory. *Journal of Electrical Systems, 20*(3), 116-123. https://doi.org/10.52783/jes.2364
26. Durlik, I., Miller, T., Kostecka, E., & Tuński, T. (2024). Artificial intelligence in maritime transportation: A comprehensive review of safety and risk management applications. *Applied Sciences, 14*(8), 8420. https://doi.org/10.3390/app14188420
27. Dutch Safety Board (DSB). (2025). *DSB official website.* https://www.onderzoeksraad.nl/en
28. European Maritime Safety Agency (EMSA). (2024). *Annual overview of marine casualties and incidents 2024.* https://www.emsa.europa.eu/publications/item/5352-annual-overview-of-marine-casualties-and-incidents-2024.html

29. European Maritime Safety Agency (EMSA). (2024). *European Marine Casualty Information Platform (EMCIP): Guide for public users. Version 4.1.0, February 2024. Lisbon*. Dostupno na: https://www.emsa.europa.eu/publications/reports.html
30. Fan, S., Zhang, J., Blanco-Davis, E., Yang, Z., & Yan, X. (2020). Maritime accident prevention strategy formulation from a human factor perspective using Bayesian networks and TOPSIS. *Ocean Engineering, 210*, 107544. https://doi.org/10.1016/j.oceaneng.2020.107544
31. Feng, Y., Wang, X., Chen, Q., Yang, Z., Wang, J., Li, H., Xia, G., & Liu, Z. (2024). Prediction of the severity of marine accidents using improved machine learning. *Transportation Research Part E: Logistics and Transportation Review, 188*, 103647. https://doi.org/10.1016/j.tre.2024.103647
32. Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications. https://www.scirp.org/reference/referencespapers?referenceid=3504991
33. Frančić, V. (2012). *Model maritimnog ustroja pomorsko-prometnog toka na prilaznim plovnim putovima* (Doktorska disertacija). Pomorski fakultet, Sveučilište u Rijeci.
34. Frénay, B., & Verleysen, M. (2014). Classification in the presence of label noise: A survey. *IEEE Transactions on Neural Networks and Learning Systems, 25*(5), 845–869. https://doi.org/10.1109/TNNLS.2013.2292894
35. Garson, G. D. (1991). *Interpreting neural network connection weights*. *Artificial Intelligence Expert, 6*(4), 46–51. https://www.scirp.org/reference/referencespapers?referenceid=1275048
36. Georgoulis, G., & Nikitakos, N. (2021). *Reporting of Near Misses: A Proactive Safety Tool in Maritime Operations*.
37. Georgoulis, G., & Nikitakos, N. (2019). *The importance of reporting all the occurred near misses on board: The seafarers' perception*. TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation, 13(3), 657-662. https://doi.org/10.12716/1001.13.03.24
38. Global Integrated Shipping Information System (GISIS). (2025). *International Maritime Organization.* https://gisis.imo.org
39. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org
40. Gül, S., Kabak, Ö., & Topcu, Y. I. (2020). Coordinating the ISM Code and OHSAS procedures to improve the occupational health and safety at sea. *Multicriteria Decision Analysis.* https://doi.org/10.1002/mcda.1710
41. Hänninen, M., Valdez Banda, O. A., & Kujala, P. (2014). Bayesian network model of maritime safety management. *Expert Systems with Applications, 41*(17), 7837–7846. https://doi.org/10.1016/j.eswa.2014.06.029
42. Hasanspahić, N., Frančić, V., Rudan, I., & Maglić, L. (2019). Analysis of navigation safety regarding tankers in narrow waterways. *Journal of Maritime & Transportation Science, 55*, 201–217. https://doi.org/10.18048/2018.00.13
43. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

44. Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415-425. https://doi.org/10.1109/72.991427
45. Hsu, W.-K. K., Chen, J.-W., Huynh, N. T., & Lin, Y.-Y. (2022). Risk assessment of navigation safety for ferries. *Journal of Marine Science and Engineering, 10*(7), Article 700. https://doi.org/10.3390/jmse10070700
46. International Maritime Organization (IMO). (2010). *Degree of risk evaluation. SN.1/Circ.296.*
47. International Maritime Organization (IMO). (2018). *Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process (MSC-MEPC.2/Circ.12/Rev.2).* https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/MSC-MEPC%202-Circ%2012-Rev%202.pdf (International Maritime Organization)
48. International Maritime Organization. (1997). *Interim guidelines for the application of Formal Safety Assessment (FSA) for use in the IMO rule-making process* (MSC/Circ.829, MEPC/Circ.335). https://wwwcdn.imo.org/localresources/en/OurWork/HumanElement/Documents/MSC-MEPC.2-Circ.12-Rev.2%20-%20Revised%20Guidelines%20For%20Formal%20Safety%20Assessment%20(Fsa)For%20Use%20In%20The%20Imo%20Rule-Making%20Proces...%20(Secretariat).pdf https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/MSC.Circ.1023-MEPC.Circ.392.pdf
49. International Maritime Organization. (2003). *Convention on the International Regulations for Preventing Collisions at Sea (COLREGs)*. https://www.imo.org/en/About/Conventions/Pages/COLREG.aspx
50. International Maritime Organization. (2008). *Guidance on Near-Miss Reporting (MSC-MEPC.7/Circ.7)*. https://www.register-iri.com/wp-content/uploads/MSC-MEPC.7-Circ.7.pdf
51. International Maritime Organization. (2018). *Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process (MSC-MEPC.2/Circ.12/Rev.2).* London: IMO.
52. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (2nd ed.). Springer. https://www.karlin.mff.cuni.cz/~pesta/NMFM334/StatLearning/Book2nd/ISLRv2_website.pdf
53. Jiang, M., Lu, J., Yang, Z., & Li, J. (2020). Risk analysis of maritime accidents along the main route of the Maritime Silk Road: A Bayesian network approach. *Maritime Policy & Management, 47*(6), 815–832. https://doi.org/10.1080/03088839.2020.1730010
54. Jiang, L., Huang, D., Liu, M., & Yang, W. (2020). Beyond synthetic noise: Deep learning on controlled noisy labels. Proceedings of the 37th International Conference on Machine Learning, 119, 4804–4815. https://proceedings.mlr.press/v119/jiang20c.html

55. Kassambara, A. (2018). *Machine Learning Essentials: Practical Guide in R*. STHDA. https://www.datanovia.com/en/wp-content/uploads/dn-tutorials/book-preview/machine-learning-essentials_preview.pdf
56. Kearney, M. W. (2017). Cramér's V. In M. R. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. n107-n107). SAGE Publications. https://doi.org/10.4135/9781483381411.n107
57. Khan, R. U., Yin, J., Asad, M., Alomayri, T., & Jameel, M. (2025). Predicting seaport disruptions from natural hazards using automated machine learning. *Marine Pollution Bulletin, 221*, 118477. https://doi.org/10.1016/j.marpolbul.2025.118477
58. Kim, Y.-J., Lee, J.-S., Pititto, A., Falco, L., Lee, M.-S., Yoon, K.-K., & Cho, I.-S. (2022). Maritime traffic evaluation using spatial-temporal density analysis based on big AIS data. *Applied Sciences, 12*(21), 11246. https://doi.org/10.3390/app122111246
59. Knapp, S., & van de Velden, M. (2024). Exploration of machine learning methods for maritime risk predictions. *Maritime Policy & Management, 51*(7), 1443-1473. https://doi.org/10.1080/03088839.2023.2209788
60. Kristiansen, S. (2005). *Maritime transportation: Safety management and risk analysis.* Oxford: Elsevier Butterworth-Heinemann.
61. Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer. https://doi.org/10.1007/978-1-4614-6849-3
62. Lan, H., Wang, S., & Zhang, W. (2024). Predicting types of human-related maritime accidents with explanations using selective ensemble learning and SHAP method. *Heliyon, 10*(9), e30046. https://www.cell.com/heliyon/fulltext/S2405-8440(24)06077-8
63. Latt, N. Z. (2024). Mitigating the risk of ship accidents with an integrated approach to maritime safety management. *Maritime Technology and Society, 3*(2), 72-79. https://doi.org/10.62012/mp.v3i2.35385
64. Li, B., Lu, J., Lu, H., & Li, J. (2021). Predicting maritime accident consequence scenarios for emergency response decisions using optimization-based decision tree approach. *Maritime Policy & Management, 48*(1), 1–23. https://doi.org/10.1080/03088839.2021.1959074
65. Li, H., Ren, X., & Yang, Z. (2023). Data-driven Bayesian network for risk analysis of global maritime accidents. *Reliability Engineering & System Safety, 230*, 108938. https://doi.org/10.1016/j.ress.2022.108938
66. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22. https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf
67. Lin, Y., Li, X., & Yuen, K. F. (2025). Machine learning applications for risk assessment in maritime transport: Current status and future directions. *Engineering Applications of Artificial Intelligence, 155*, 110959. https://doi.org/10.1016/j.engappai.2025.110959
68. Liu, K., Yu, Q., Yuan, Z., Yang, Z., & Shu, Y. (2021). A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches. *Ocean & Coastal Management, 213*, 105859. https://doi.org/10.1016/j.ocecoaman.2021.105859

69. Liu, R., Yan, R., Wang, R., Hanaoka, S., & Wang, S. (2025). Semi-supervised learning for multi-target regression in ship risk prediction. *Maritime Policy & Management*. Advance online publication. https://doi.org/10.1080/03088839.2025.2503828
70. Lu, J., Su, W., Jiang, M., & Ji, Y. (2022). Severity prediction and risk assessment for non-traditional safety events in sea lanes based on a random forest approach. *Ocean & Coastal Management, 225*(3), 106202. https://doi.org/10.1016/j.ocecoaman.2022.106202
71. Ma, Q., Du, X., Zhang, M., Wang, H., Lang, X., & Mao, W. (2024). A spatial-temporal attention method for the prediction of multi-ship time headways using AIS data. *Ocean Engineering, 311*(Part 2), 118927. https://doi.org/10.1016/j.oceanengl.2024.118927
72. Marine Accident and Incident Investigation Committee of the Republic of Cyprus (MAIC). (2025). *Home page.* https://www.gov.cy/mtcw/en/marine-accident-and-incident-investigation-committee-maic
73. Marine Accident Investigation Branch (MAIB). (2025). *MAIB Data Portal: Power BI Dashboard.* https://maps.dft.gov.uk/maib-data-portal/web-pages/pbi_dashboard.html
74. Mishra, D., Tripathi, S. M., Chaurasia, A., & Chaurasia, P. K. (2025). A review on ensemble learning methods: Machine learning approach. *International Journal of Research Publication and Reviews, 6*(2), 3795-3803. https://doi.org/10.55248/gengpi.6.0225.0971
75. Mohović, Đ. (2010). *Ocjena prihvatljivosti pomorskih plovidbenih rizika* (Doktorska disertacija). Sveučilište u Rijeci, Pomorski fakultet.
76. Mohović, R. (2002). *Model manevriranja brodom u ograničenim plovnim područjima u funkciji sigurnosti i zaštite morskog okoliša* (Doktorska disertacija). Sveučilište u Rijeci, Pomorski fakultet.
77. Mok, I., D'Agostini, E., & Ryoo, D.-K. (2023). A validation study of ISM Code's continual effectiveness through a multilateral comparative analysis of maritime accidents in Korean waters. *The Journal of Navigation, 76*(1), 77-90. https://doi.org/10.1017/S0373463322000571
78. Mou, F., Fan, Z., Li, X., Wang, L., & Li, X. (2024). A method for clustering and analyzing vessel sailing routes efficiently from AIS data using traffic density images. *Journal of Marine Science and Engineering, 12*(1), 75. https://doi.org/10.3390/jmse12010075
79. Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69-71. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576830/
80. Munim, Z. H., Sørli, M. A., Kim, H., & Alon, I. (2024). Predicting maritime accident risk using Automated Machine Learning. *Reliability Engineering & System Safety,* 248, 110148. https://doi.org/10.1016/j.ress.2024.110148
81. Musulin, M., Kero, G., Matijašević, L., & Peša, T. (2024a). Analysis of statistical correlation of maritime accidents and navigation safety activities in the Adriatic Sea. *Transportation Research Procedia, 76*, 8. https://doi.org/10.1016/j.trpro.2025.02.040
82. Musulin, M., Kero, G., Musulin, H.N., & Brčić, D. (2023). A port entry risk assessment model based on Bayesian networks and elements of the e-navigation concept. *Pomorstvo: Scientific Journal of Maritime Research, 37*(1), 140-150.

https://doi.org/10.31217/p.37.1.11
83. Musulin, M., Mihanović, L., Balić, K., & Musulin, H. (2024b). The impact of container ship trim on fuel consumption and navigation safety. *Journal of Marine Science and Engineering, 12*(9), 1658. https://doi.org/10.3390/jmse12091658
84. National Ocean Service - NOAA, National Oceanic and Atmospheric Administration. (2025). *Beaufortova ljestvica* Dostupno na: https://www.weather.gov/mfl/beaufort
85. National Transportation Safety Board (NTSB). (2025). *CAROL Basic Search.* https://data.ntsb.gov/carol-main-public/basic-search
86. Omarzai, F. (2024, July 20). *Random forest in-depth*. Medium. https://medium.com/@fraidoonomarzai99/random-forest-in-depth-f0556817c40b (Medium)
87. Onyshchenko, S., Bychkovsky, Y., Melnyk, O., & Onishchenko, O. (2024). A model for assessing shipping safety within project-orientated risk management based on human element. *Scientific Journal of Silesian University of Technology. Series Transport, 123*, 319-334. https://doi.org/10.20858/sjsutst.2024.123.16
88. Park, J., & Jeong, J.-S. (2021). An estimation of ship collision risk based on relevance vector machine. *Journal of Marine Science and Engineering, 9*(5), 538. https://doi.org/10.3390/jmse9050538
89. Perkovič, M., Gucma, L., & Feuerstack, S. (2024). Maritime security and risk assessments. *Journal of Marine Science and Engineering, 12*(6), 988. https://doi.org/10.3390/jmse12060988
90. PIANC. (1992). *Capability of ship manoeuvring simulation models for approach channels and fairways in harbours* (Report of Working Group No. 20 of Permanent Technical Committee II, Supplement to PIANC Bulletin No. 77).
91. Portillo Juan, N., Negro Valdecantos, V., & Troch, P. (2025). Advancing artificial intelligence in ocean and maritime engineering: Trends, progress, and future directions. *Ocean Engineering, 339* (Part 1), 122077. https://doi.org/10.1016/j.oceanengl.2025.122077
92. ProtectUK. (2025). *Risk management basics.* https://www.protectuk.police.uk/risk-management-basics
93. Raschka, S. (2018). *Model evaluation, model selection, and algorithm selection in machine learning* (Technical report). University of Wisconsin-Madison, Department of Statistics. https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/model-eval-selection.pdf
94. Rawson, A., & Brito, M. (2022). A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis. *Transport Reviews*. https://doi.org/10.1080/01441647.2022.2036864
95. Rawson, A., Brito, M., & Sabeur, Z. (2021). Spatial modeling of maritime risk using machine learning. *Risk Analysis, 41*(12), 2372–2388. https://doi.org/10.1111/risa.13866
96. Rawson, A., Brito, M., & Sabeur, Z. (2022). Spatial modeling of maritime risk using machine learning. *Risk Analysis, 42*(10), 2153-2169. https://doi.org/10.1111/risa.13866
97. Rawson, A., Brito, M., Sabeur, Z., & Tran-Thanh, L. (2021). From conventional to machine learning methods for maritime risk assessment. *TransNav: The International*

Journal on Marine Navigation and Safety of Sea Transportation, 15(4), 757-764. https://doi.org/10.12716/1001.15.04.06

98. Republika Hrvatska. (2021). *Zakon o plovidbi i lukama unutarnjih voda (NN 144/21).* Narodne novine.
99. Rifkin, R., & Klautau, A. (2004). In defense of one-vs-all classification. *Journal of Machine Learning Research*, 5, 101-141. http://www.jmlr.org/papers/volume5/rifkin04a/rifkin04a.pdf
100. Roy, A., & Kasiviswanathan, K. S. (2023). Exploring the potential of data-driven models for streamflow simulation in Himalayan region. In *Developments in Environmental Science* (Vol. 14, pp. 193–215). Elsevier. https://doi.org/10.1016/B978-0-443-18640-0.00010-9
101. RStudio Team. (2024). *RStudio: Integrated development environment for R.* https://www.rstudio.com/
102. Safety Investigation Authority of Finland (SAIF). (2025). *Marine accident investigation.* https://www.turvallisuustutkinta.fi/eRoyn/index/tutkintaselostukset.html
103. Safety4Sea. (2022, September 12). *Why reporting near misses is vital*. https://safety4sea.com/cm-why-reporting-near-misses-is-vital/
104. Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763-1768. https://pubmed.ncbi.nlm.nih.gov/29481436
105. Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press. https://mcube.lab.nycu.edu.tw/~cfung/docs/books/scholkopf2002learning_with_kernels.pdf
106. scikit-learn (n.d.). Classification metrics. https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics
107. Spire. (2024). *How RightShip leverages machine learning and AI for proactive risk management* [Case study]. https://spire.com/case-study/maritime/how-rightship-leverages-machine-learning-and-ai-for-proactive-risk-management/
108. Swedish Accident Investigation Authority (SAIA). (2025). *Statens haverikommission.* https://www.shk.se/en/
109. Thakur, A. S., Alex, L., & Nighojkar, A. (2025). Artificial intelligence in maritime anomaly detection: A decadal bibliometric analysis (2014-2024). *Journal of The Institution of Engineers (India): Series C, 106*(2), 665-689. https://doi.org/10.1007/s40032-025-01169-w
110. Transport Accident Investigation Commission (TAIC). (2025). *Investigations.* https://www.taic.org.nz/inquiries
111. Transport Safety Investigation Center - Marine accident investigation (UEIM). (2025). https://denizcilikistatistikleri.uab.gov.tr/turk-bogazlari-gemi-gecis-istatistikleri
112. Transportation Safety Board of Canada (TSBC). (2025). *Marine investigation reports.* https://www.tsb.gc.ca/eng/rapports-reports/marine/index.html
113. Ung, S. T. (2021). Navigation risk estimation using a modified Bayesian network modeling—A case study in Taiwan. *Reliability Engineering & System Safety, 213*, 107777. https://doi.org/10.1016/j.ress.2021.107777

114. United Nations. (1982). *United Nations Convention on the Law of the Sea* (Art. 76). https://www.un.org/depts/los/convention_agreements/texts/unclos/unclos_e.pdf
115. Uyanik, G. K., & Guler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*, 106, 234-240. https://www.sciencedirect.com/science/article/pii/S1877042813046429
116. Vapnik, V. N. (1998). *Statistical learning theory*. Wiley. https://www.scirp.org/reference/referencespapers?referenceid=84632
117. Waltz, M., Paulig, N., & Okhrin, O. (2025). 2-level reinforcement learning for ships on inland waterways: Path planning and following. *Expert Systems with Applications, 274*, 126933. https://doi.org/10.1016/j.eswa.2025.126933
118. Wang, H., Liu, Z., Wang, X., Graham, T., & Wang, J. (2021). An analysis of factors affecting the severity of marine accidents. *Reliability Engineering & System Safety, 210*, 107513. https://doi.org/10.1016/j.ress.2021.107513
119. Wang, L., & Yang, Z. (2018). Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China. *Reliability Engineering & System Safety, 180*, 277–289. https://doi.org/10.1016/j.ress.2018.07.021
120. Wang, L., Wang, J., Hua, Y., Shi, W., Yang, Z., & Sha, M. (2025). Machine learning approaches for identifying substandard ships in port state control inspections with imbalanced data. *Ocean Engineering, 334*, 121614. https://doi.org/10.1016/j.oceanengl.2025.121614
121. White, D. R. (2003). *A student's guide to statistics for analysis of cross tabulations*. University of California, Irvine. https://www.researchgate.net/publication/245086225_A_Student's_Guide_to_Statistics_for_Analysis_of_Cross_Tabulations
122. Wolpert, D. H. (1992). *Stacked generalization*. Neural Networks, 5(2), 241-259. https://doi.org/10.1016/S0893-6080(05)80023-1
123. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1
124. Wu, L., Xu, Y., Wang, Q., Wang, F., & Xu, Z. (2016). Mapping global shipping density from AIS data. The Journal of Navigation, 70(1), 67-81. https://doi.org/10.1017/S0373463316000345
125. Xue, J., Yang, P., Li, Q., Song, Y., van Gelder, P. H. A. J. M., Papadimitriou, E., & Hu, H. (2025). Machine learning in maritime safety for autonomous shipping: A bibliometric review and future trends. *Journal of Marine Science and Engineering, 13*(4), 746. https://doi.org/10.3390/jmse13040746
126. Yang, J., Han, S., & Chen, Y. (2023). Prediction of traffic accident severity based on random forest. *Journal of Advanced Transportation*, Article 7641472. https://doi.org/10.1155/2023/7641472
127. Yi, M.-S., Lee, B.-K., & Park, J.-S. (2025). Data-driven analysis of causes and risk assessment of marine container losses: Development of a predictive model using machine learning and statistical approaches. *Journal of Marine Science and Engineering, 13*(3), 420. https://doi.org/10.3390/jmse13030420
128. Yildiz, S., Uğurlu, Ö., Loughney, S., Wang, J., & Tonoğlu, F. (2022). Spatial and statistical analysis of operational conditions influencing accident formation in narrow

waterways: A case study of Istanbul Strait and Dover Strait. *Ocean Engineering, 265*, 112692. https://doi.org/10.1016/j.oceanengl.2022.112692

129. Yuan, X., Wang, J., Zhao, G., & Wang, H. (2024). Comprehensive study on optimizing inland waterway vessel routes using AIS data. *Journal of Marine Science and Engineering, 12*(10), 1775. https://doi.org/10.3390/jmse12101775
130. Zhang, L., Zhu, Y., Valdez Banda, O. A., Du, L., Gan, L., & Li, X. (2024). Navigational risk assessment of inland waters based on bi-directional PSO-LSTM algorithm and ship maneuvering characteristics. *Ocean Engineering, 310* (Part 1), 118628. https://doi.org/10.1016/j.oceanengl.2024.118628
131. Zhang, M., Kujala, P., & Hirdaris, S. (2022). A machine learning method for the evaluation of ship grounding risk in real operational conditions. *Reliability Engineering & System Safety, 226*, 108697. https://doi.org/10.1016/j.ress.2022.108697
132. Zhang, W., Ma, X., & Zhang, Y. (2025a). Research on navigation risk assessment index system of intelligent ships. *Ocean Engineering, 322*, 120435. https://doi.org/10.1016/j.oceanengl.2025.120435
133. Zhang, Y., Li, G., Zhu, J., & Cheng, X. (2025b). Machine learning-based binary classification models for low ice-class vessels navigation risk assessment. *Journal of Marine Science and Engineering, 13*(8), 1408. https://doi.org/10.3390/jmse13081408
134. Zhang, Z., Hu, Q., & Yin, J. (2025c). Maritime-accident-induced environmental pollution and economic loss analysis using an interpretable data-driven method. *Sustainability, 17*(7), 3023. https://doi.org/10.3390/su17073023
135. Zhou, Z.-H. (2021). Ensemble learning. In *Machine Learning* (pp. 181-210). Springer. https://doi.org/10.1007/978-981-15-1967-3_8
136. Zhu, F., & Liu, L. (2013). Research on a method for analysis of ship traffic density in harbor water area based on GIS. In *IEEE Conference Anthology* (pp. 1–4). https://doi.org/10.1109/ANTHOLOGY.2013.6785051

POPIS KRATICA

Kratica	Naziv na engleskom jeziku	Naziv na hrvatskom jeziku
ACC	*Accuracy*	Točnost
ADASYN	*Adaptive Synthetic Sampling*	Adaptivno sintetičko uzorkovanje
AI	*Artificial Intelligence*	Umjetna inteligencija
AIS	*Automatic Identification System*	Sustav za automatsku identifikaciju
ALARP	*As Low As Reasonably Practicable*	Na razumno ostvarivo najnižoj razini
AN	*Actual Negative*	Stvarno negativno
ANN	*Artificial Neural Network*	Umjetna neuronska mreža
AP	*Actual Positive*	Stvarno pozitivno
ARM	*Aviation Risk Management*	Upravljanje rizicima u zračnom prometu
ARPA	*Automatic Radar Plotting Aid*	Sustav automatskog radarskog plotiranja
ATSB	*Australian Transport Safety Bureau*	Australski ured za sigurnost prometa
AUC	*Area Under the Curve*	Površina ispod krivulje
AutoML	*Automated Machine Learning*	Automatizirano strojno učenje
BN	*Bayesian Network*	Bayesova mreža
BO	*Bayesian Optimization*	Bayesova optimizacija
BSU	*Federal Bureau for Maritime Casualty Investigation (Germany)*	Savezni ured za istraživanje pomorskih nezgoda (Njemačka)
COLREG	*International Regulations for Preventing Collisions at Sea*	Međunarodni propisi za izbjegavanje sudara na moru
CSL	*Cost-sensitive Learning*	Ponderirani modeli učenja
DL	*Deep Learning*	Duboko učenje
DMAIB	*Danish Maritime Accident Investigation Board*	Danski odbor za istraživanje pomorskih nezgoda
DNN	*Deep Neural Network*	Duboka neuronska mreža
DOR	*Diagnostic Odds Ratio*	Dijagnostički omjer
DSB	*Dutch Safety Board*	Nizozemski odbor za sigurnost
ECDIS	*Electronic Chart Display and Information System*	Informacijski sustav i prikaz elektroničkih karata
EMCIP	*European Marine Casualty Information Platform*	Europska platforma za informacije o pomorskim nezgodama
EMODnet	*European Marine Observation and Data Network*	Europska mreža za nadzor i prikupljanja podataka o moru
EMSA	*European Maritime Safety Agency*	Europska agencija za pomorsku sigurnost
ErrorRate	*Error Rate*	Stopa pogreške
ETA	*Event Tree Analysis*	Analiza stabla događaja
F1	*F1 Score*	F1-mjera

FDR	*False Discovery Rate*	Stopa lažno pozitivnih predikcija
FMEA	*Failure Modes and Effects Analysis*	Analiza mogućih kvarova i njihovih posljedica
FN	*False Negative*	Lažno negativni
FNR	*False Negative Rate*	Stopa lažno negativnih slučajeva
FOI	*Freedom of Information*	Pravo na pristup informacijama
FOR	*False Omission Rate*	Stopa lažno negativnih predikcija
FP	*False Positive*	Lažno pozitivni
FPR	*False Positive Rate*	Stopa lažno pozitivnih slučajeva
FSA	*Formal Safety Assessment*	Formalna sigurnosna procjena
FTA	*Fault Tree Analysis*	Analiza stabla grešaka
GA	*Genetic Algorithms*	Genetski algoritmi
GIS	*Geographic Information System*	Geografski informacijski sustav
GISIS	*Global Integrated Shipping Information System*	Globalni integrirani informacijski sustav za brodarstvo
GT (BT)	*Gross tonagge*	Bruto tonaža
HAZOP	*Hazard and Operability Study*	Studija opasnosti i operabilnosti
HRMF	*High-resolution Meteorological Forecasts*	Meteorološke prognoze visoke rezolucije
HV	*Hydrographic Variables*	Hidrografske varijable
IMO	*International Maritime Organisation*	Međunarodna pomorska organizacija
ISM	*International Safety Management Code*	Međunarodni kodeks za upravljanje sigurnošću
IVI	*Interactive Visualization Interfaces*	Interaktivna vizualizacijska sučelja
LOG	*Logistics*	Logistika
LR^-	*Negative Likelihood Ratio*	Omjer negativne vjerojatnosti
LR^+	*Positive Likelihood Ratio*	Omjer pozitivne vjerojatnosti
LSTM	*Long Short-Term Memory*	Dugoročna kratkoročna memorija
LSTO	*Long Short-Term Optimization*	Dugoročna kratkoročna optimizacija
MAIB	*Marine Accident Investigation Branch*	Ured za istragu pomorskih nezgoda pri Ujedinjenom Kraljevstvu
MAIC	*Marine Accident and Incident Investigation Committee*	Odbor za pomorske nezgode i incidente (Cipar)
MCS	*Monte Carlo Simulation*	Monte Carlo simulacija
ML	*Machine Learning*	Strojno učenje
N/A	*Not Applicable*	Neprimjenjivo
NA	*Not Available ili Missing Value*	Nedostajuća vrijednost
NM	*Nautical Mile*	Nautička milja
NN	*Neural Networks*	Neuronske mreže
NPV	*Negative Predictive Value*	Negativna prediktivna vrijednost

NTSB	*National Transportation Safety Board*	Nacionalni odbor za sigurnost prometa (SAD)
OHSAS	*Occupational Health and Safety Assessment Series*	Serija normi za procjenu zaštite i sigurnosti na radu
ORM	*Operational Risk Management*	Operativno upravljanje rizicima
PN	*Predicted Negative*	Predviđeno kao negativno
PP	*Predicted Positive*	Predviđeno kao pozitivno
PPV	*Positive Predictive Value*	Pozitivna prediktivna vrijednost / Preciznost
PSO	*Particle Swarm Optimization*	Optimizacija rojem čestica
RBF	*Radial Base Function*	Radijalna bazna funkcija
RCA	*Root Cause Analysis*	Analiza osnovnog uzroka
REG	*Red Ensign Group*	Skupina zastava pod jurisdikcijom UK-a
ReLU	*Rectified Linear Unit*	Rektificirana linearna jedinica
RL	*Reinforcement Learning*	Učenje s pojačanjem
RM	*Random Forest*	Slučajne šume
SAIA	*Swedish Accident Investigation Authority*	Švedska agencija za istraživanje nezgoda
SAR	*Search and Rescue*	Traganje i spašavanje
SBRM	*Scenario-based Risk Models*	Scenarijski modeli procjene rizika
SHK	*Swedish Accident Investigation Authority*	Švedska agencija za istraživanje nezgoda
SIAF	*Safety Investigation Authority of Finland*	Agencija za istraživanje nezgoda (Finska)
SMOTE	*Synthetic Minority Over-sampling Technique*	Sintetičko naduzorkovanje manjinske klase
SO	*Satellite Observations*	Satelitska osmatranja
STAT	*Spatio-temporal Analytical Techniques*	Prostorno-vremenske analitičke tehnike
ST-DBSCAN	*Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise*	Prostorno-vremensko grupiranje gustoće podataka s uključenim šumom
SVM	*Support Vector Machine*	Metoda potpornih vektora
TAIC	*Transport Accident Investigation Commission*	Komisija za istragu prometnih nezgoda (Novi Zeland)
TN	*True Negative*	Istiniti negativni
TNR	*True Negative Rate*	Stvarna negativna stopa / Specifičnost
TP	*True Positive*	Istiniti pozitivni
TPE	*Tree-structured Parzen Estimators*	Parzenovi procjenitelji strukturirani u obliku stabla

TPR	*True Positive Rate*	Stvarna pozitivna stopa / Osjetljivost
TSB	*Transportation Safety Board of Canada*	Odbor za sigurnost prometa (Kanada)
UEIM	*Ministry of Transport and Infrastructure (Turkey)*	Turski centar za istraživanje sigurnosti u prometu
UNCLOS	*United Nations Convention on the Law of the Sea*	Konvencija Ujedinjenih naroda o pravu mora
VTS	*Vessel Traffic Service*	Služba nadzora i upravljanja pomorskim prometom
XAI	*Explainable Artificial Intelligence*	Objašnjiva umjetna inteligencija
XGBoost	*Extreme Gradient Boosting*	Ekstremna gradijentna pojačana metoda

POPIS SLIKA

Slika 1. Shema metodološkog pristupa istraživanja....6
Slika 2. Grafički prikaz hiperravnine i potpornih vektora u SVM klasifikaciji....21
Slika 3. Općenita struktura modela slučajnih šuma....22
Slika 4. Grafički prikaz arhitekture višeslojne neuronske mreže....24
Slika 5. Grafički prikaz metodološkog okvira istraživanja....35
Slika 6. Globalna distribucija pomorskih nezgoda....52
Slika 7. Pomorske nezgode prema zastavi broda....58
Slika 8. Deset najsnažnijih Spearmanovih korelacija među ordinalnim varijablama....89
Slika 9. Deset najsnažnih Pearsonovih korelacija između ordinalnih varijabli....90
Slika 10. Usporedba Spearmanovih i Pearsonovih korelacija za deset najsnažnijih odnosa između ordinalnih varijabli....91
Slika 11. Deset najsnažnijih korelacija između nominalnih varijabli prema Cramérovu V koeficijentu....92
Slika 12. Deset najsnažnijih korelacija između ordinalnih i nominalnih varijabli prema Cramérovu V koeficijentu....93
Slika 13. Najvažnije varijable u klasifikaciji pomorskih nezgoda kod modela potpornih vektora....99
Slika 14. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda metodom potpornih vektora....102
Slika 15. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih modelom potpornih vektora i pripadajućih indeksa posljedica....103
Slika 16. Najvažnije varijable u klasifikaciji pomorskih nezgoda kod modela slučajnih šuma....107
Slika 17. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda metodom slučajnih šuma....108
Slika 18. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih modelom slučajnih šuma i pripadajućih indeksa posljedica....109
Slika 19. Struktura neuronske mreže....112
Slika 20. Najvažnije varijable u klasifikaciji pomorskih nezgoda kod modela neuronskih mreža....114
Slika 21. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda metodom neuronskih mreža....116
Slika 22. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih modelom neuronskih mreža i pripadajućih indeksa posljedica....117
Slika 23. Usporedni prikaz točno klasificiranih slučajeva po glavnim događajima i modelima (SVM, RF, i NN)....119
Slika 24. Usporedni prikaz točno klasificiranih slučajeva po svim klasama i svim modelima (SVM, RF, NN, ansambl i slojeviti)....124
Slika 25. Usporedni prikaz metrika uspješnosti klasifikacijskih modela....126
Slika 26. Poredak modela prema ukupnoj točnosti....126
Slika 27. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda primjenom ansambl modela....128
Slika 28. Utjecaj najvažnijih varijabli na vjerojatnost pomorskih nezgoda primjenom slojevitog modela....129
Slika 29. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih primjenom ansambl modela i pripadajućih indeksa posljedica....130
Slika 30. Matrica rizika glavnih događaja na temelju vjerojatnosti procijenjenih primjenom slojevitog modela i pripadajućih indeksa posljedica....131
Slika 31. Usporedba ukupnog doprinosa riziku (ΣR_i) po modelima strojnog učenja....132
Slika 32. Toplinska matrica raspodjele rizika po glavnim događajima i modelima....133
Slika 33. Toplinska matrica raspodjele vjerojatnosti po glavnim događajima i modelima....134
Slika 34. Odnos klasifikacijske točnosti (ACC) i ukupnog doprinosa riziku (ΣR_i)....134

Slika 35. Usporedba doprinosa riziku po skupinama varijabli i modelima 135

POPIS TABLICA

Tablica 1. Ključna istraživanja o procjeni rizika u obalnoj plovidbi analizirana u ovom radu 28
Tablica 2. Pregled metodološkog okvira analize podataka istraživanja 31
Tablica 3. Matrica rizika 37
Tablica 4. Kriteriji klasifikacije razine učestalosti i posljedica prema FSA metodologiji 37
Tablica 5. Opis svih varijabli ulaznog skupa podataka 44
Tablica 6. Strukturne karakteristike ulaznog skupa podataka 46
Tablica 7. Značenja vrijednosti varijable *glavni događaj* (*main_event*) 48
Tablica 8. Značenje i broj pojavljivanja vrijednosti varijable *baza podataka* (*database*) 48
Tablica 9. Klasifikacija i broj pojavljivanja varijable *ozbiljnost nezgode* (*occurrence*) 49
Tablica 10. Vrijednosti i broj pojavljivanja varijable *godina izvještavanja* (*year_reported*) 50
Tablica 11. Broj prijavljenih pomorskih nezgoda prema *mjesecu* (*month*) 50
Tablica 12. Broj prijavljenih pomorskih nezgoda prema *državi izvjestiteljici* (*state_reporting*) 51
Tablica 13. Klasifikacija i broj pojavljivanja varijable *prirodno osvjetljenje* (*natural_light)* 52
Tablica 14. Douglas-ova ljestvica i broj pojavljivanja varijable *stanja mora* (*sea_state*) 53
Tablica 15. Klasifikacija i broj pojavljivanja varijable *vidljivost* (*visibility)* 54
Tablica 16. Klasifikacija i broj pojavljivanja varijable *jačina vjetra* (*wind_force*) prema Beaufortovoj ljestvici 54
Tablica 17. Klasifikacija i broj pojavljivanja varijable *vremenski uvjeti* (*weather*) 55
Tablica 18. Klasifikacija i broj pojavljivanja varijable *mjesto pomorske nezgode* (*occurrence_location*) 55
Tablica 19. Klasifikacija i broj pojavljivanja varijable *gustoća prometa* (*density*) 56
Tablica 20. Vrijednosti varijable *područje nezgode* (*location_area*) 57
Tablica 21. Vrijednosti varijable *zastava broda* (*vessel_flag*) 59
Tablica 22. Klasifikacija i broj pojavljivanja varijable *kategorija broda* (*vessel_category*) 60
Tablica 23. Klasifikacija i broj pojavljivanja varijable *bruto tonaža* (*gt*) 60
Tablica 24. Klasifikacija i broj pojavljivanja varijable *tip broda* (*vessel_type*) 61
Tablica 25. Klasifikacija i broj pojavljivanja varijable *duljina* broda (*length*) 62
Tablica 26. Klasifikacija i broj pojavljivanja varijable *širina* broda (*width)* 62
Tablica 27. Klasifikacija i broj pojavljivanja varijable *vrsta štete* (*damage*) 63
Tablica 28. Klasifikacija i broj pojavljivanja varijable *onečišćenje* (*pollution*) 63
Tablica 29. Klasifikacija i broj pojavljivanja varijable *vrsta materijala broda* (*material*) 64
Tablica 30. Klasifikacija i broj pojavljivanja varijable *starost broda* (*ship_age*) 64
Tablica 31. Klasifikacija i broj pojavljivanja varijable *potonuće broda* (*vessel_sink)* 64
Tablica 32. Klasifikacija i broj pojavljivanja varijable *broj članova posade* (*crew*) 65
Tablica 33. Klasifikacija i broj pojavljivanja varijable *segment plovidbe* (*voyage_segment*) 66
Tablica 34. Klasifikacija i broj pojavljivanja varijable *peljar na brodu* (*pilot*) 66
Tablica 35. Klasifikacija i broj pojavljivanja varijable *ukupno ozljeda na brodu* (*injuries_total*) 67
Tablica 36. Matrica zabune 74
Tablica 37. Korištene metrike za ocjenu klasifikacijskih modela 75
Tablica 38. Deset najsnažnijih Spearmanovih korelacija između ordinalnih varijabli 89
Tablica 39. Deset najsnažnih Pearsonovih korelacija između ordinalnih varijabli 90
Tablica 40. Deset najsnažnijih Cramérovih V koeficijenata između nominalnih varijabli 92
Tablica 41. Deset najsnažnijih Cramérovih V koeficijenata između ordinalnih i nominalnih varijabli 93
Tablica 42. Empirijska distribucija pojavnosti i ponderiranih doprinosa po glavnim događajima 95

Tablica 43. Agregirani koeficijent C_i po glavnim događajima 96
Tablica 44. Matrica zabune za model potpornih vektora 97
Tablica 45. Ključne klasifikacijske metrike modela potpornih vektora 97
Tablica 46. Varijable najsnažnijeg utjecaja u modelu potpornih vektora 98
Tablica 47. Matrica zabune za model slučajnih šuma 104
Tablica 48. Ključne klasifikacijske metrike modela slučajnih šuma 105
Tablica 49. Varijable najsnažnijeg utjecaja u modelu slučajnih šuma 106
Tablica 50. Matrica zabune za model neuronske mreže 110
Tablica 51. Ključne klasifikacijske metrike modela neuronskih mreža 111
Tablica 52. Varijable najsnažnijeg utjecaja u modelu neuronske mreže 113
Tablica 53. Pravilno klasificirani slučajevi na testnom skupu 118
Tablica 54. Analiza metrika klasifikacije 119
Tablica 55. Matrica zabune ansambl modela 121
Tablica 56. Ključne klasifikacijske metrike ansambl modela 122
Tablica 57. Matrica zabune slojevitog modela 122
Tablica 58. Ključne klasifikacijske metrike slojevitog modela 123
Tablica 59. Usporedni prikaz pravilno klasificiranih slučajeva po klasama i modelima 124
Tablica 60. Usporedni prikaz metrika uspješnosti klasifikacijskih modela 125
Tablica 61. Usporedni prikaz procijenjenih vjerojatnosti pojave glavnih događaja i izvedenih razina rizika za modele strojnog učenja 131
Tablica 62. Shema kontrolirane izmjene ciljnih vrijednosti varijable glavni događaj 136
Tablica 63. Metrike modela u odnosu na kontrolirano promijenjene ciljne oznake 136

PRILOZI

Prilog A Programski kod u R .....**Error! Bookmark not defined.**

Prilog B Tablica B1. Pregled varijabli prema tipu podataka i načinu obrade u analizi ...**Error! Bookmark not defined.**

Prilog C Tablica B2. Detaljni rezultati modela potpornih vektora: procijenjene vjerojatnosti P_i, koeficijent posljedica C_i i doprinos riziku R_i po varijablama i glavnim događajima.**Error! Bookmark not defined.**

Prilog D Tablica B3. Detaljni rezultati modela slučajnih šuma: procijenjene vjerojatnosti P_i, koeficijent posljedica C_i i doprinos riziku R_i po varijablama i glavnim događajima.....**Error! Bookmark not defined.**

Prilog E Tablica B4. Detaljni rezultati modela neuronske mreže: procijenjene vjerojatnosti P_i, koeficijent posljedica C_i i doprinos riziku R_i po varijablama i glavnim događajima.**Error! Bookmark not defined.**

Prilog F Tablica B5. Detaljni rezultati ansambl modela: procijenjene vjerojatnosti P_i, koeficijent posljedica C_i i doprinos riziku R_i po varijablama i glavnim događajima.....**Error! Bookmark not defined.**

Prilog G Tablica B6. Detaljni rezultati slojevitog modela: procijenjene vjerojatnosti P_i, koeficijent posljedica C_i i doprinos riziku R_i po varijablama i glavnim događajima.....**Error! Bookmark not defined.**